

DOCUMENT RESUME

ED 026 079

LI 000 980

By-Agenbroad, James E.; And Others

Systems Design and Pilot Operation of a Regional Center for Technical Processing for the Libraries of the New England State Universities. NELINET, New England Library Information Network. Progress Report, July 1, 1967 - March 30, 1968, Volume II, Appendices.

Inforonics, Inc., Cambridge, Mass.

Spons Agency-New England Board of Higher Education, Wellesley, Mass.

Pub Date 5 Apr 68

Contract-CLR-385

Note-169p.; Vol. I is LI 000 979.

EDRS Price MF-\$0.75 HC-\$8.55

Descriptors-Automation, Cataloging, Centralization, Information Processing, Library Acquisition, *Library Networks, *Library Technical Processes, Pilot Projects, Regional Programs, *Systems Development, *University Libraries

Identifiers-NELINET, *New England Library Information Network

Included in this volume of appendices to LI 000 979 are acquisitions flow charts; a current operations questionnaire; an algorithm for splitting the Library of Congress call number; analysis of the Machine-Readable Cataloging (MARC II) format; production problems and decisions; operating procedures for information transmittal in the New England Library Information Network; compression word coding techniques (transition distance coding, alphacheck, recursive decomposition, and Soundex); and sample cards and labels. (CC)

nfor *



LI 000980

ED026079

NELINET - NEW ENGLAND LIBRARY INFORMATION NETWORK

PROGRESS REPORT

JULY 1, 1967 - MARCH 30, 1968

SYSTEMS DESIGN AND PILOT OPERATION OF A
REGIONAL CENTER FOR TECHNICAL PROCESSING
FOR THE LIBRARIES OF THE NEW ENGLAND
STATE UNIVERSITIES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

VOLUME II - APPENDICES

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
POSITION OR POLICY.

PREPARED BY

JAMES E. AGENBROAD, LAWRENCE F. BUCKLAND, ANN T. CURRAN,
DONALD D. HODGINS, WILLIAM R. NUGENT, ROBERT H. SIMMONS

SUBMITTED TO

THE NEW ENGLAND BOARD OF HIGHER EDUCATION

FINAL REPORT

CONTRACT NO. CLR - 385

APRIL 5, 1968

08600027

TABLE OF CONTENTS

APPENDIX A - ACQUISITIONS FLOW CHARTS

APPENDIX B - CURRENT OPERATIONS QUESTIONNAIRE

APPENDIX C - AN ALGORITHM FOR SPLITTING THE LC CALL NUMBER

APPENDIX D - ANALYSIS OF THE MARC II FORMAT

APPENDIX E - PRODUCTION PROBLEMS AND DECISIONS

APPENDIX F - OPERATING PROCEDURES FOR INFORMATION TRANSMITTAL

IN THE NEW ENGLAND LIBRARY INFORMATION NETWORK

APPENDIX G - COMPRESSION WORD CODING TECHNIQUES: TRANSITION

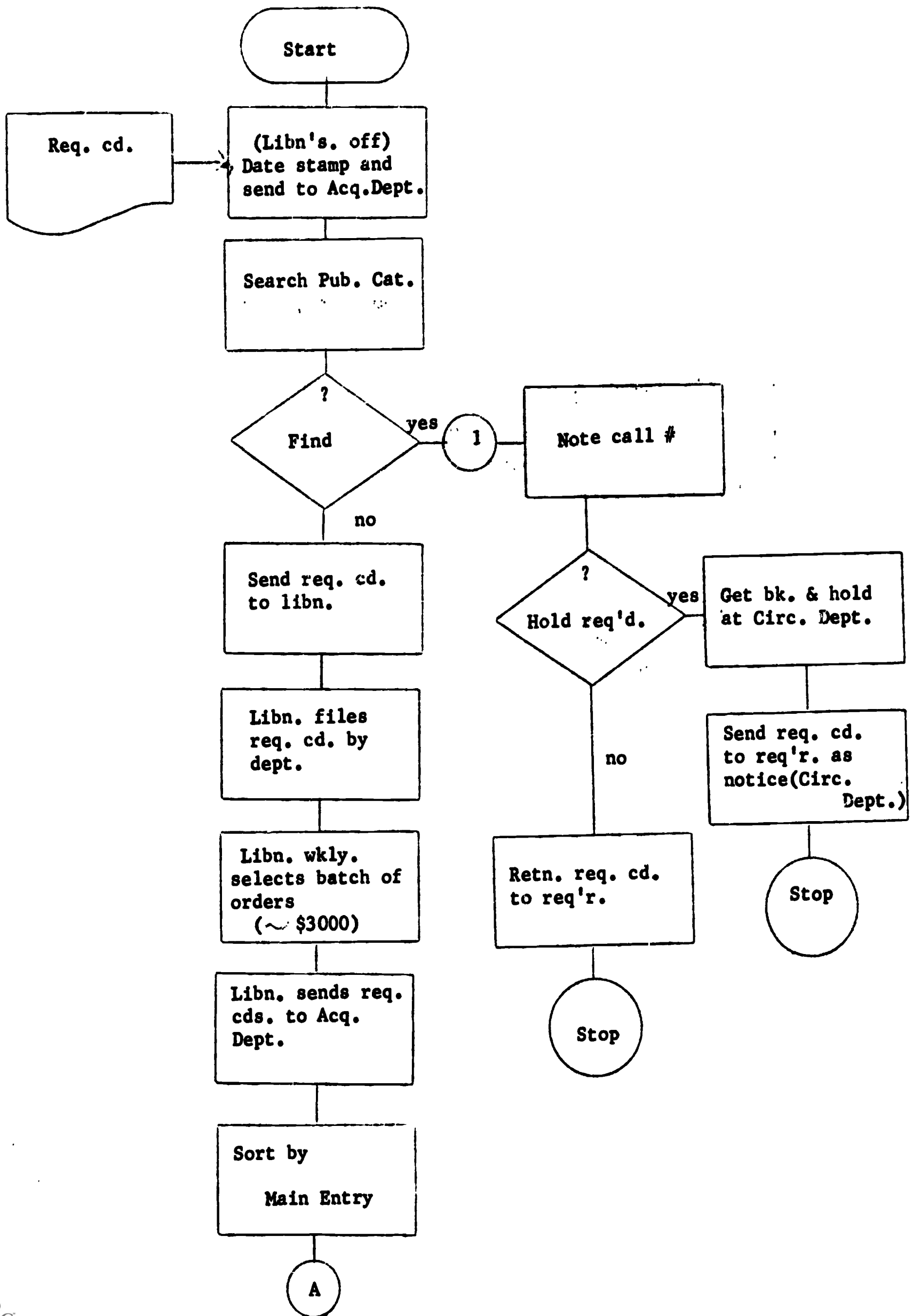
DISTANCE CODING, ALPHACHECK, RECURSIVE DECOMPOSITION

AND SOUNDEX

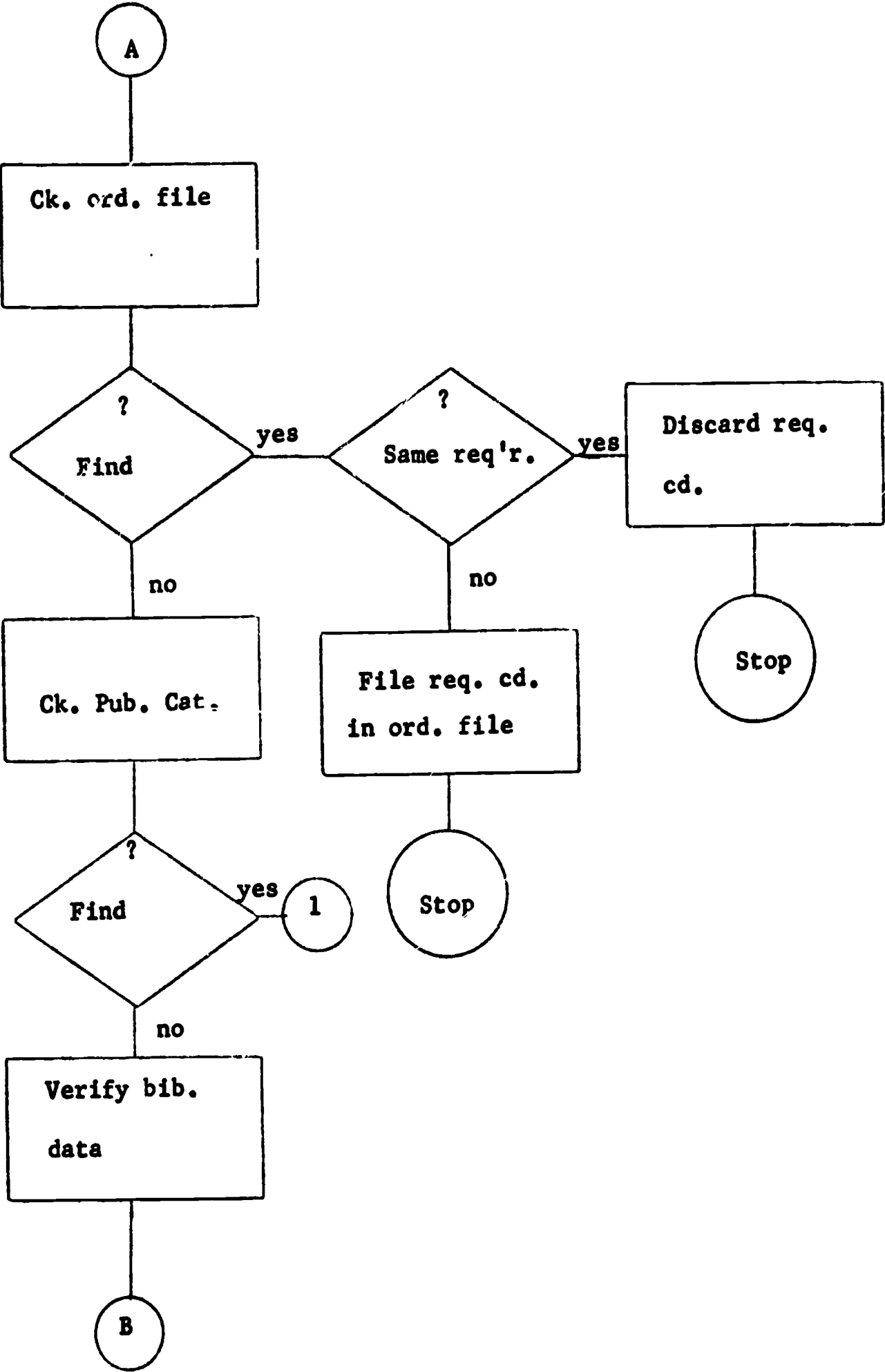
APPENDIX H - SAMPLE CARDS AND LABELS

APPENDIX A

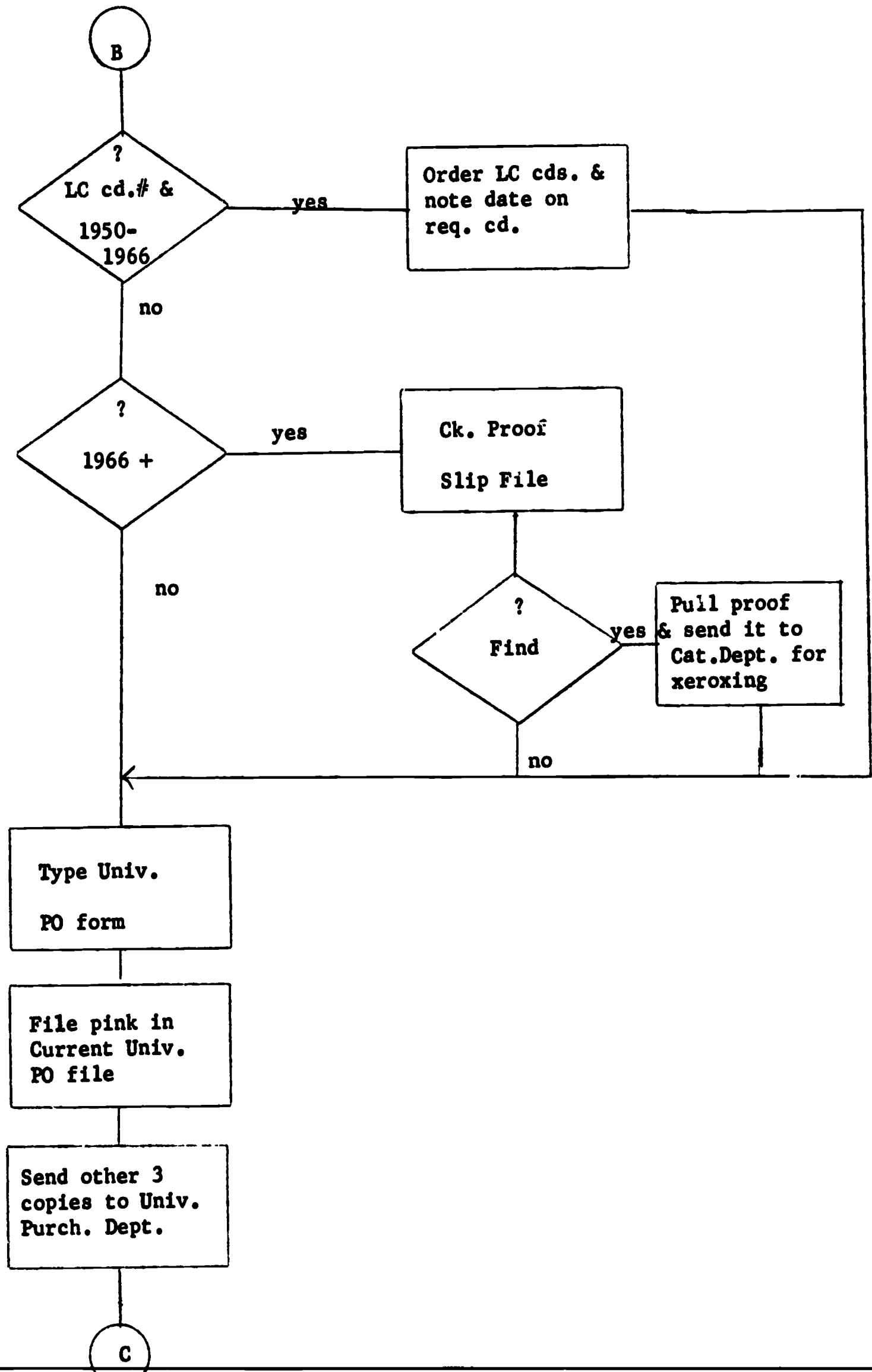
Library A--Acquisitions Department--Ordering



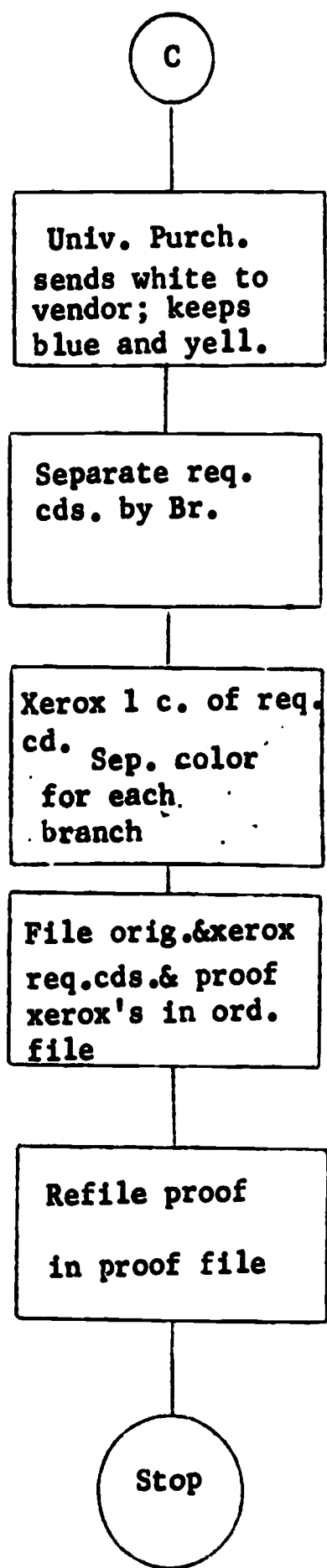
A - 2
Library A--Acquisitions Department--Ordering



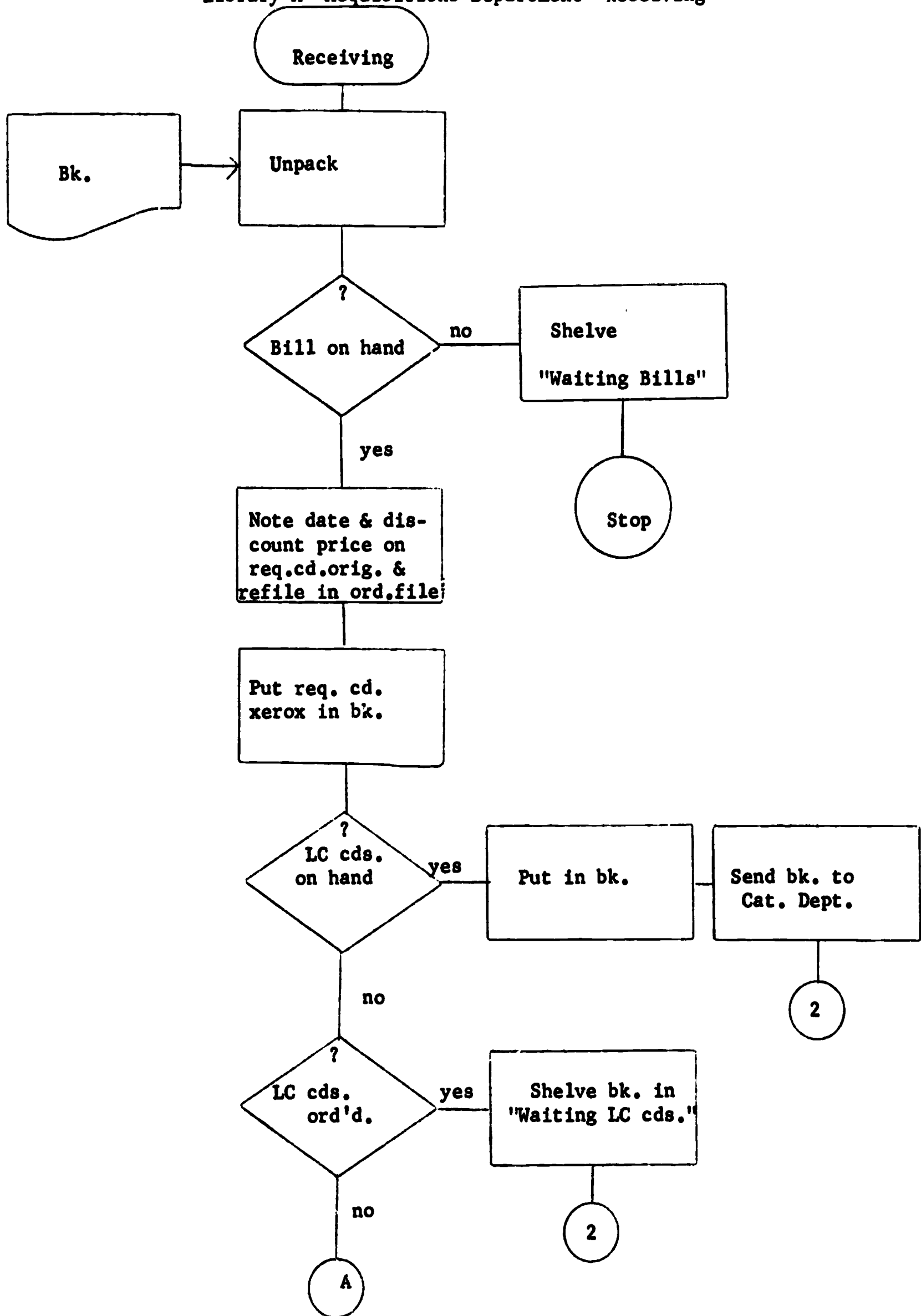
Library A--Acquisitions Department--Ordering



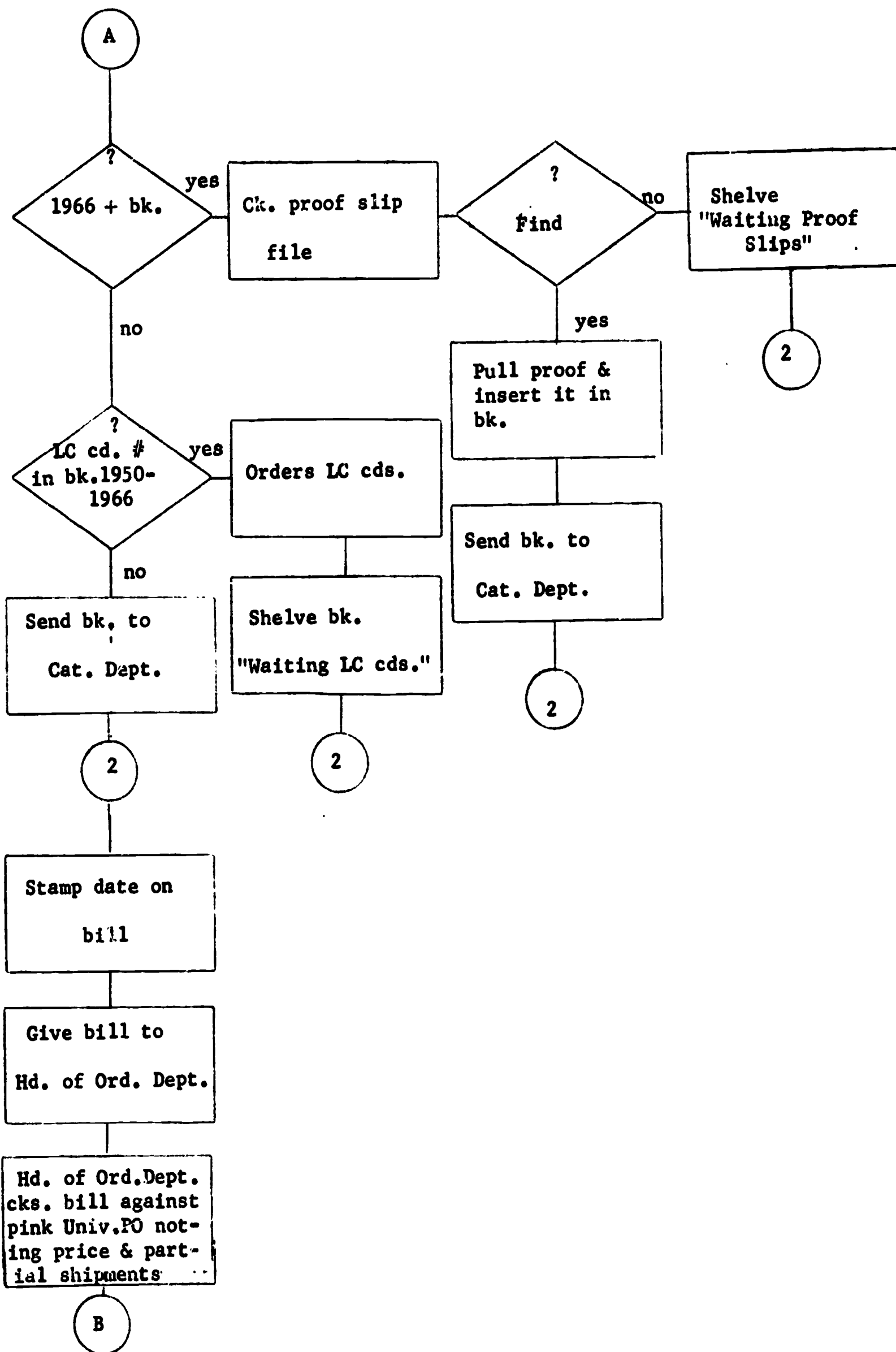
Library A--Acquisitions Department--Ordering



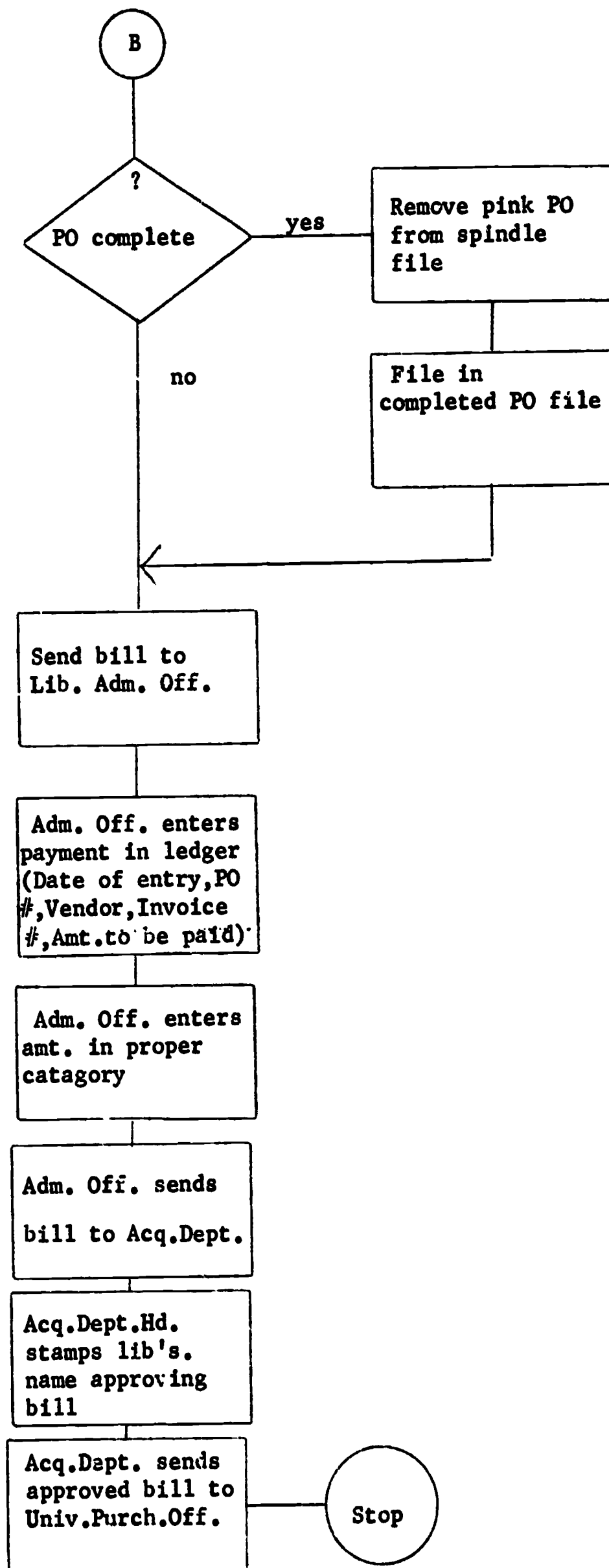
Library A--Acquisitions Department--Receiving



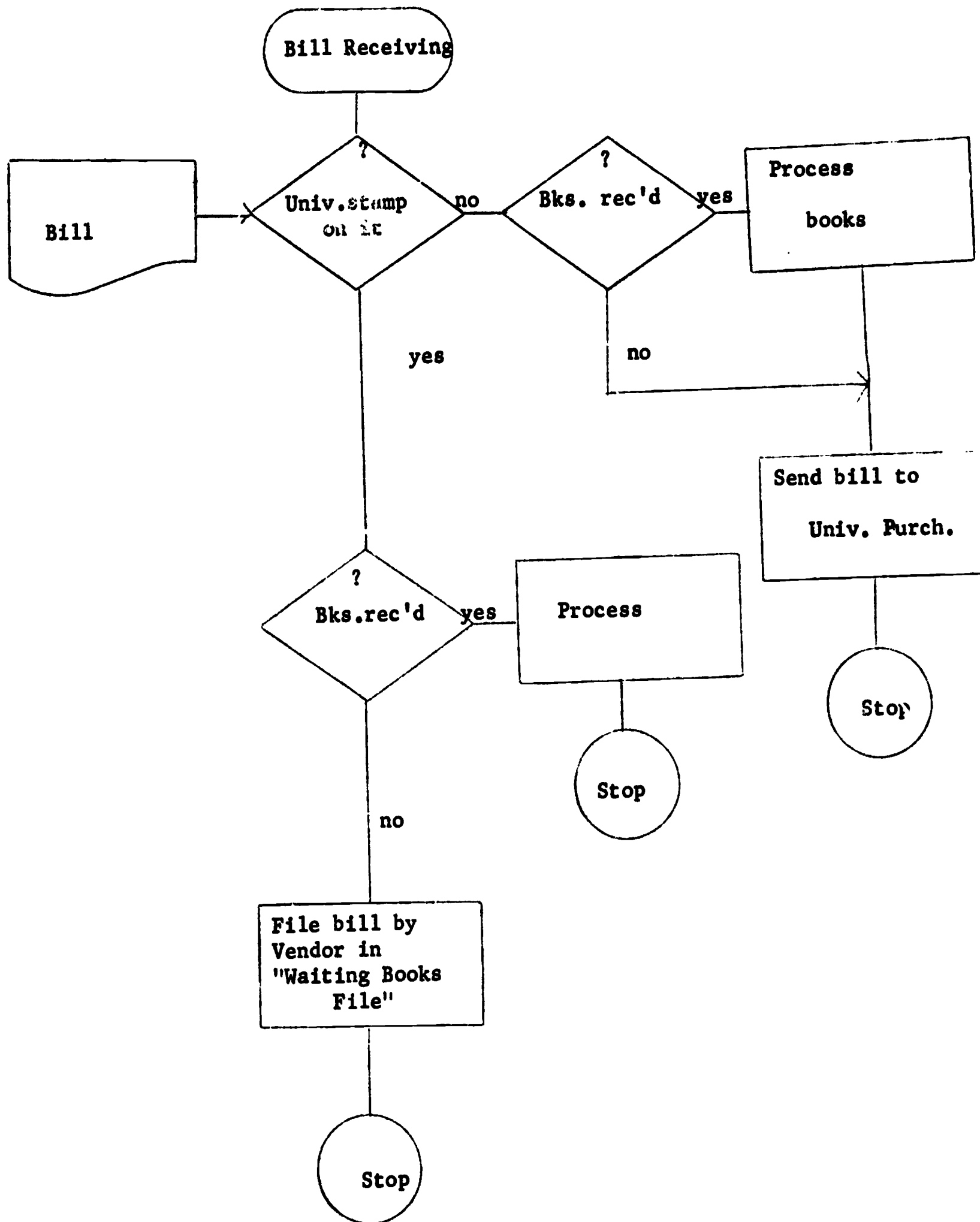
Library A--Acquisitions Department--Receiving



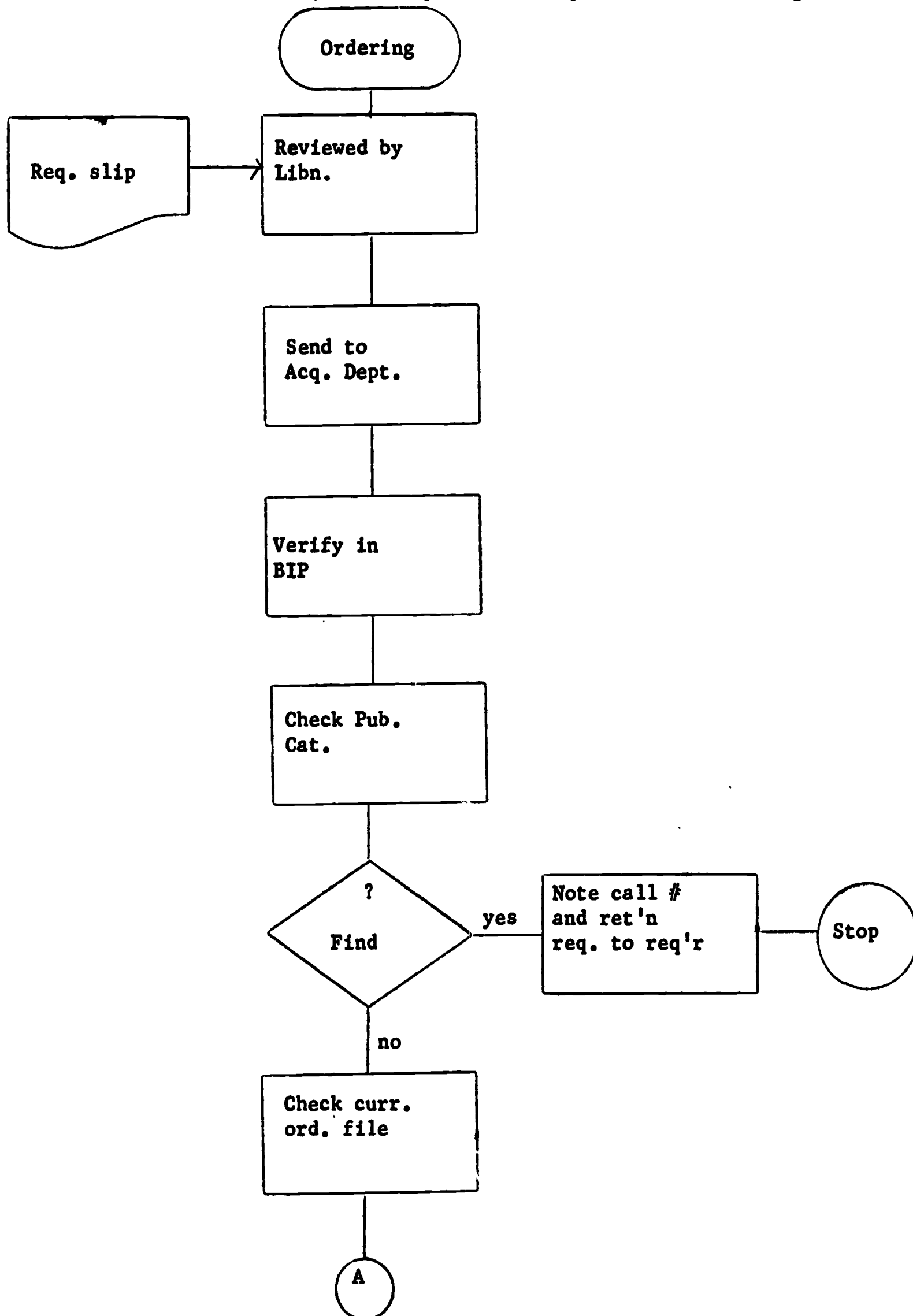
Library A--Acquisitions Department--Receiving



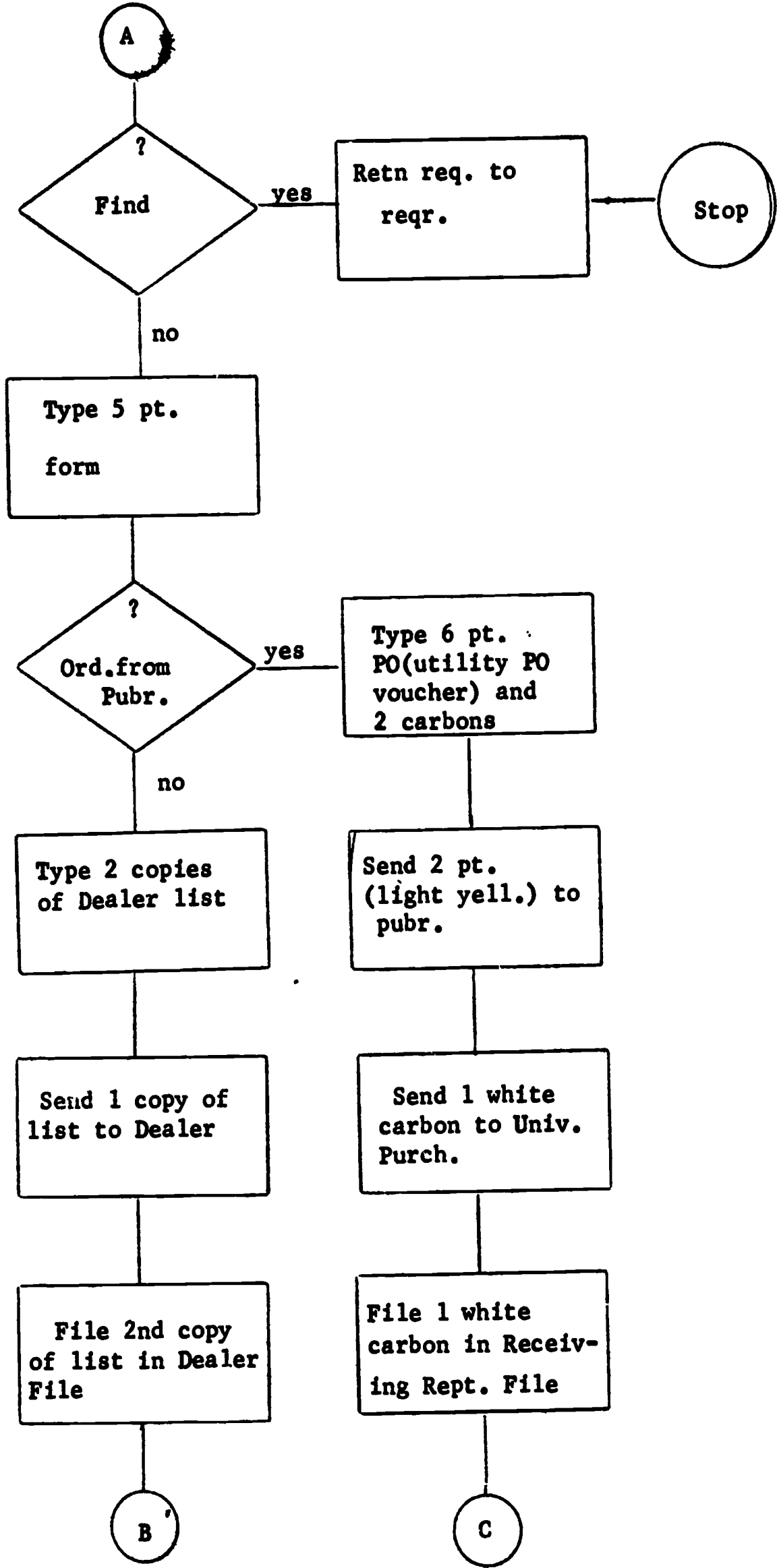
Library A--Acquisitions Department--Receiving



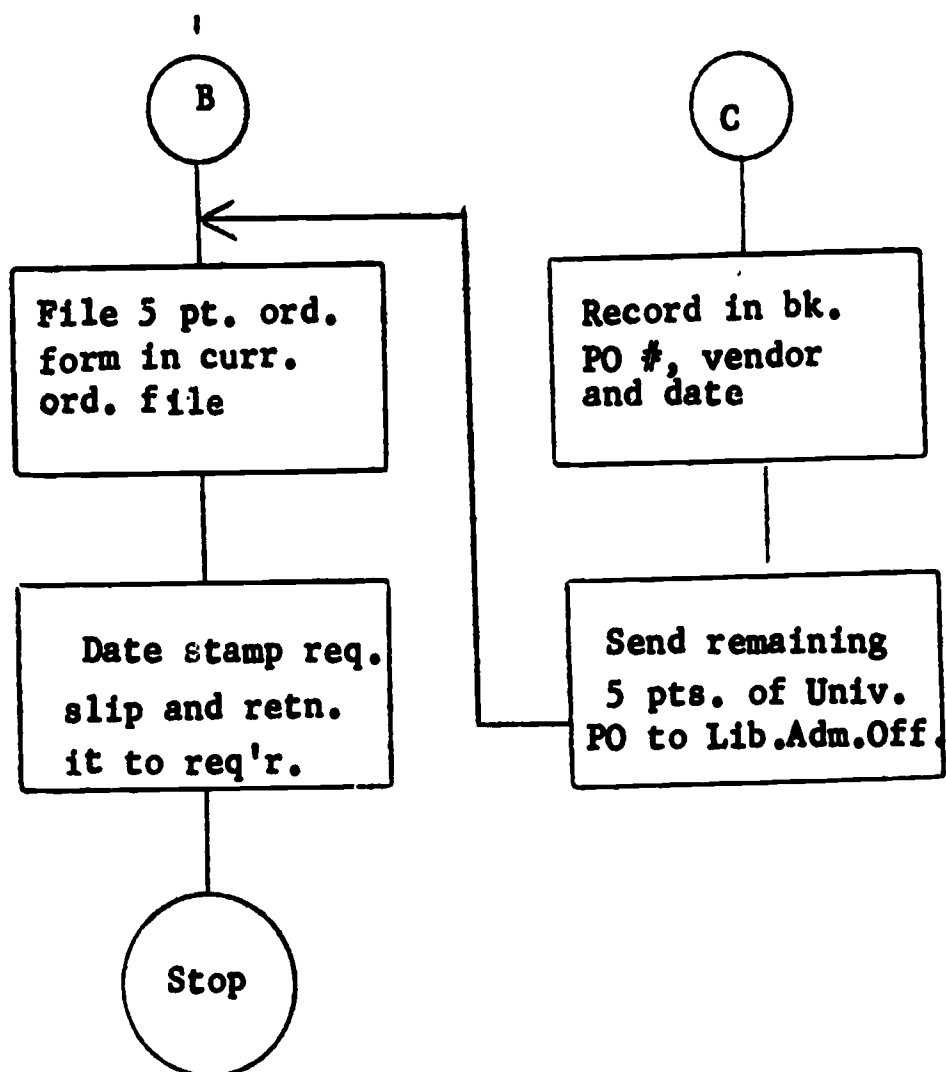
Library B -- Acquisitions Department -- Ordering



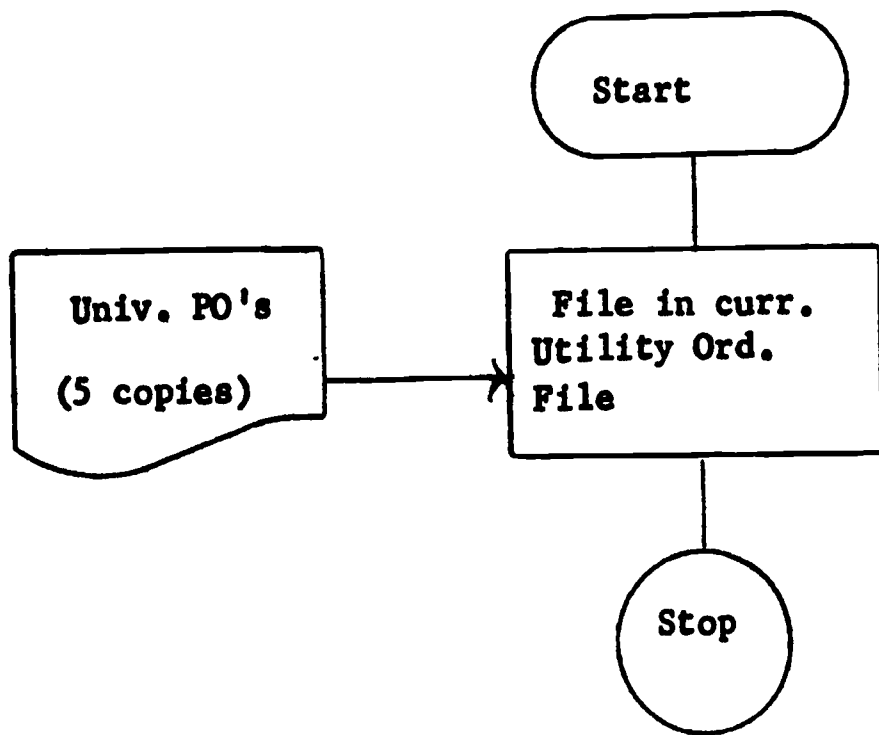
A - 10
Library B--Acquisitions Department--Ordering



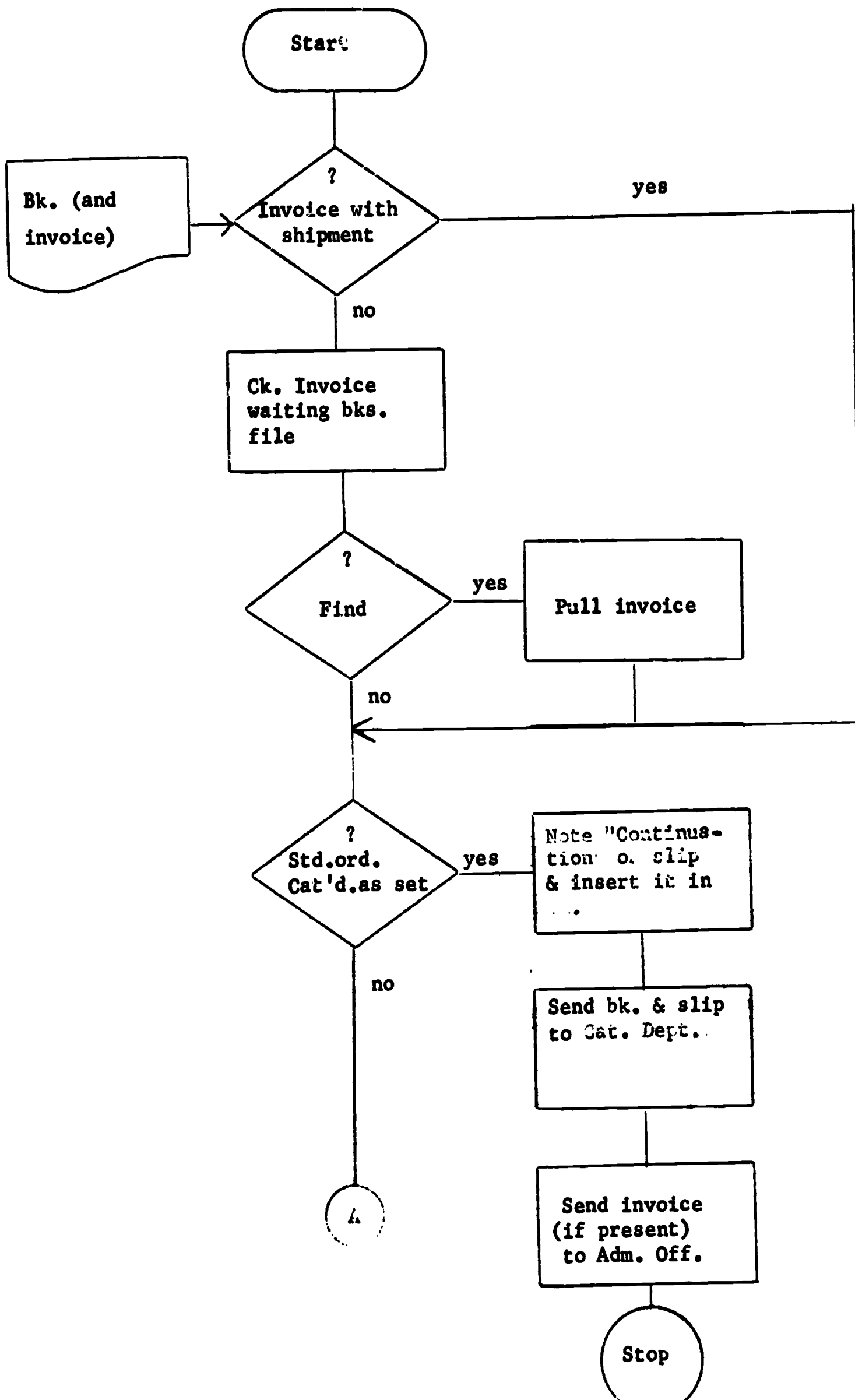
A - 11
Library B--Acquisitions Department--Ordering



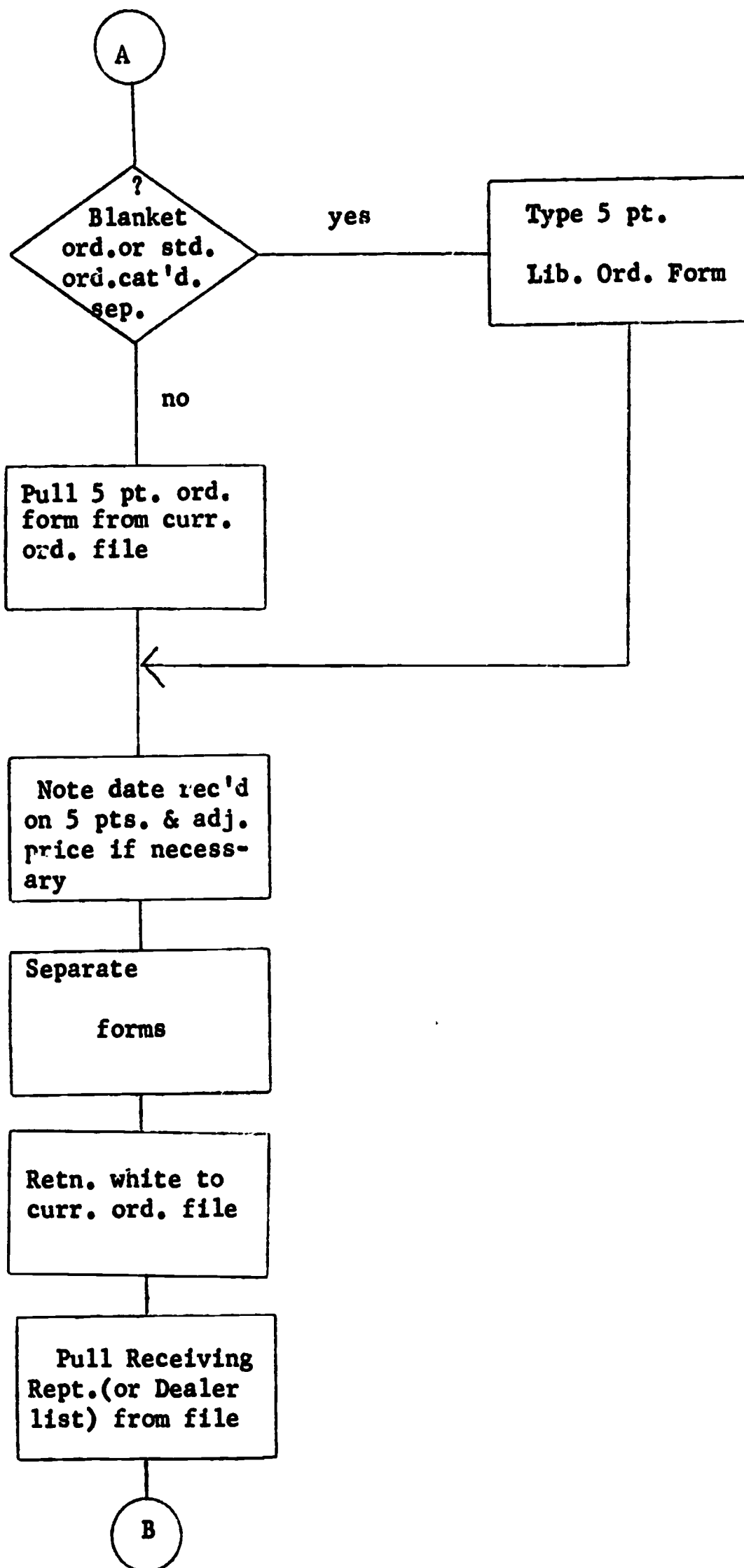
A - 12
Library B--Library Administration Office--
--Ordering--



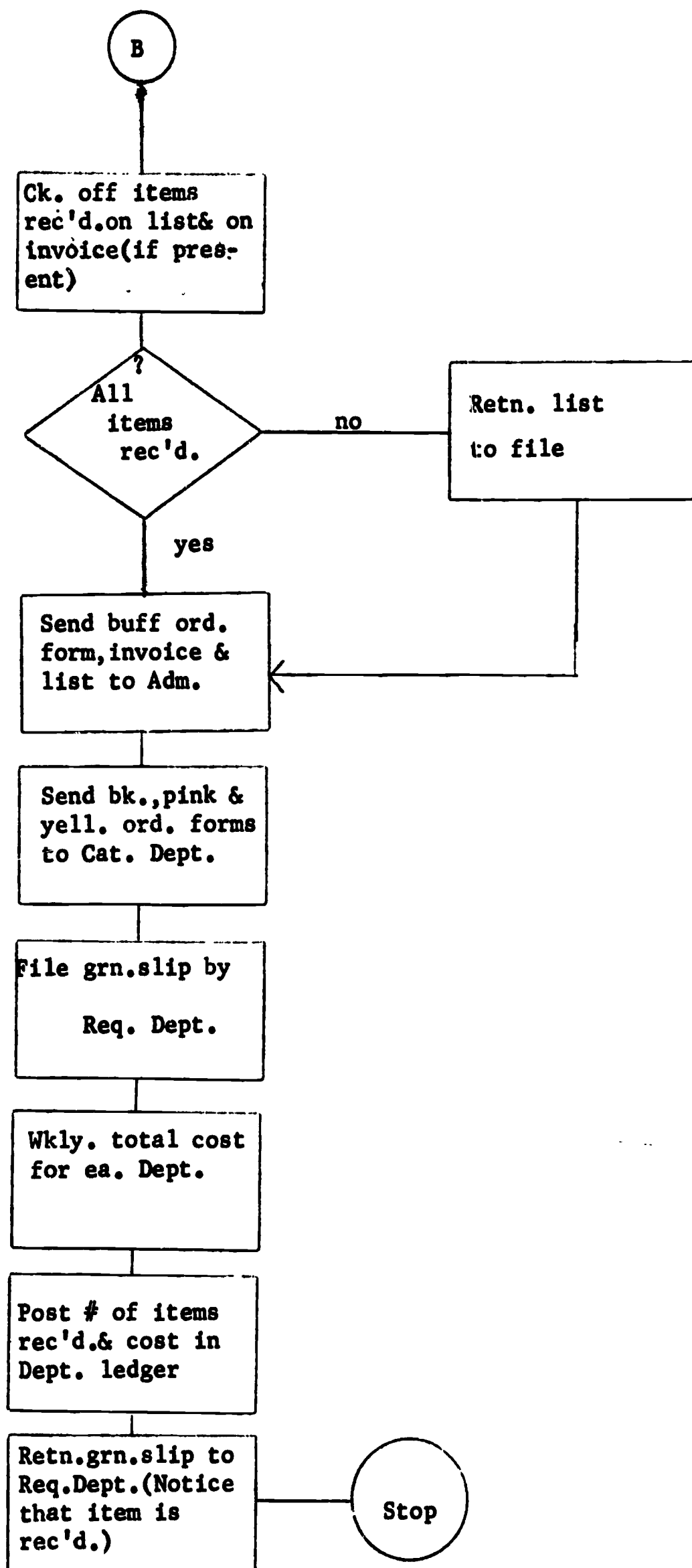
A - 13
Library B--Acquisitions Department--Receiving



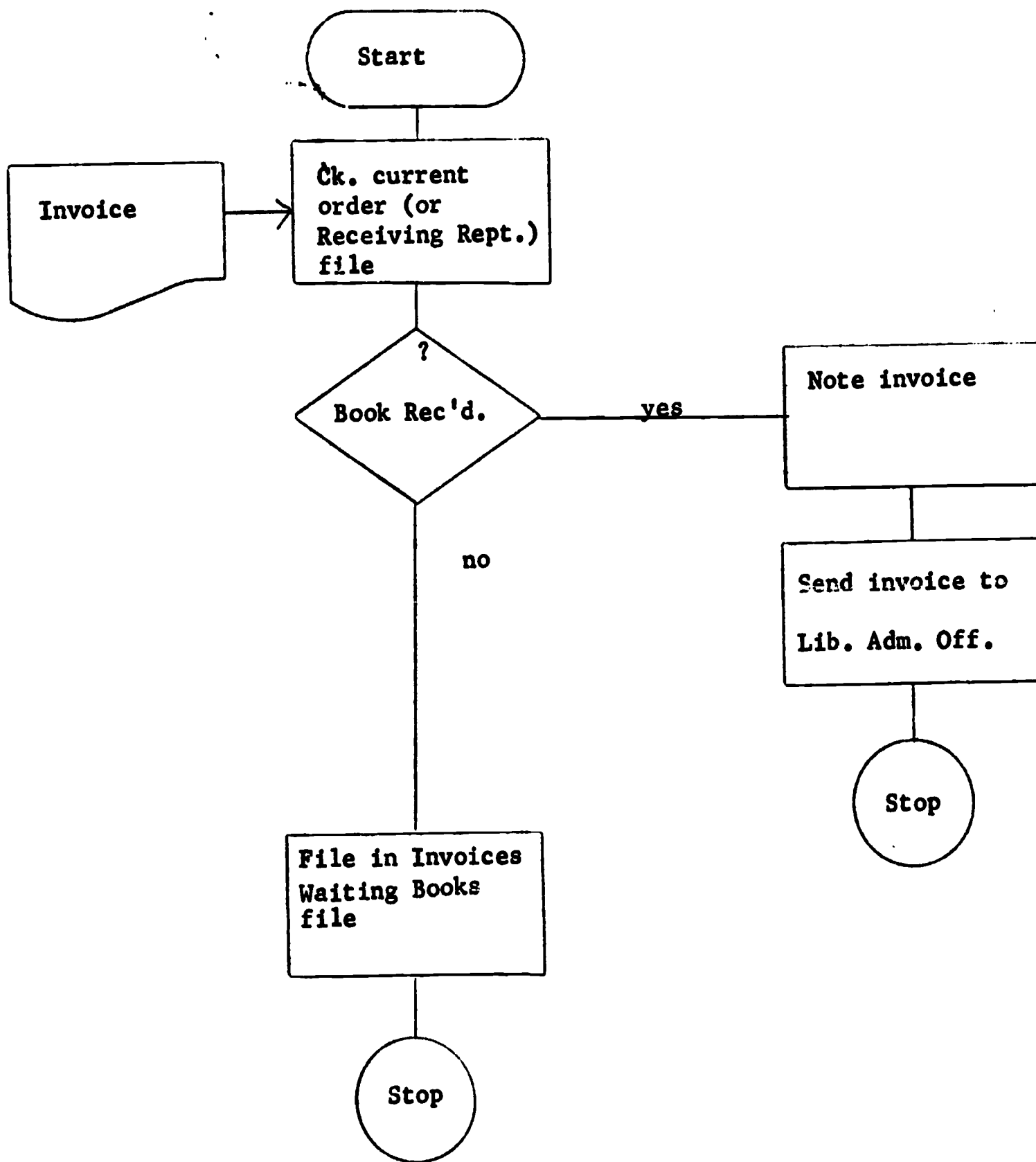
Library B--Acquisitions Department--Receiving



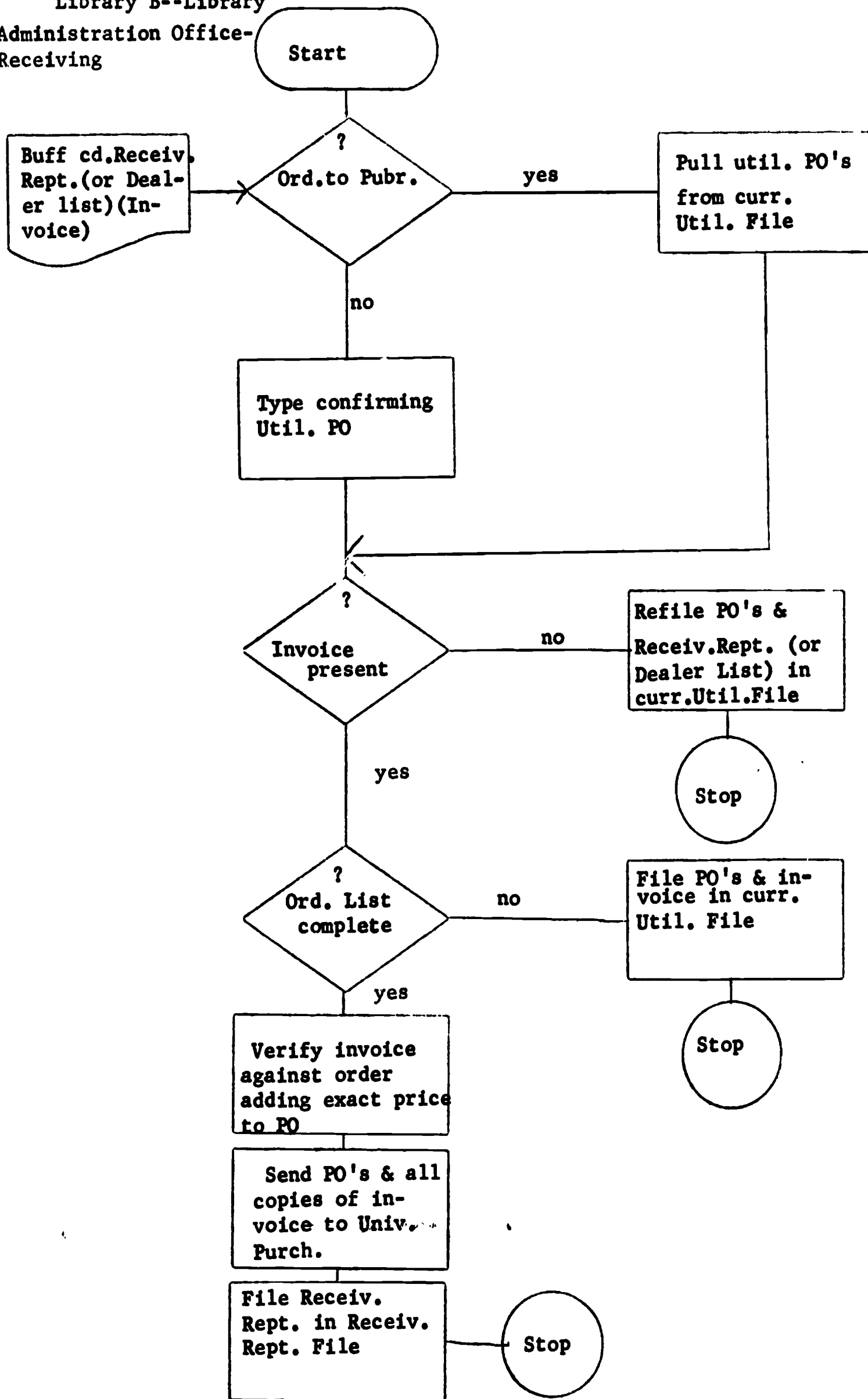
Library B--Acquisitions Department--Receiving



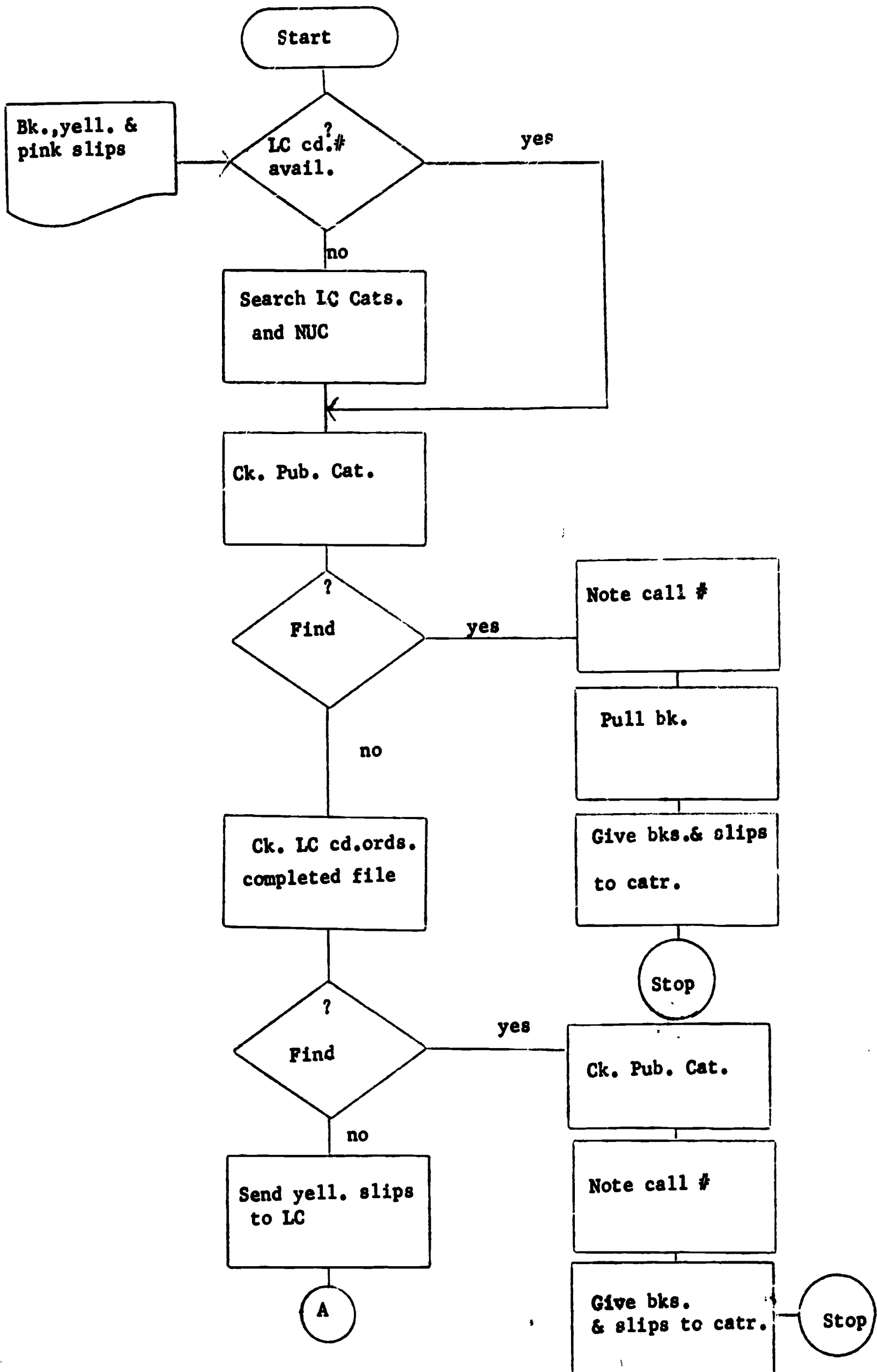
A - 16
Library B--Acquisitions Department--Receiving



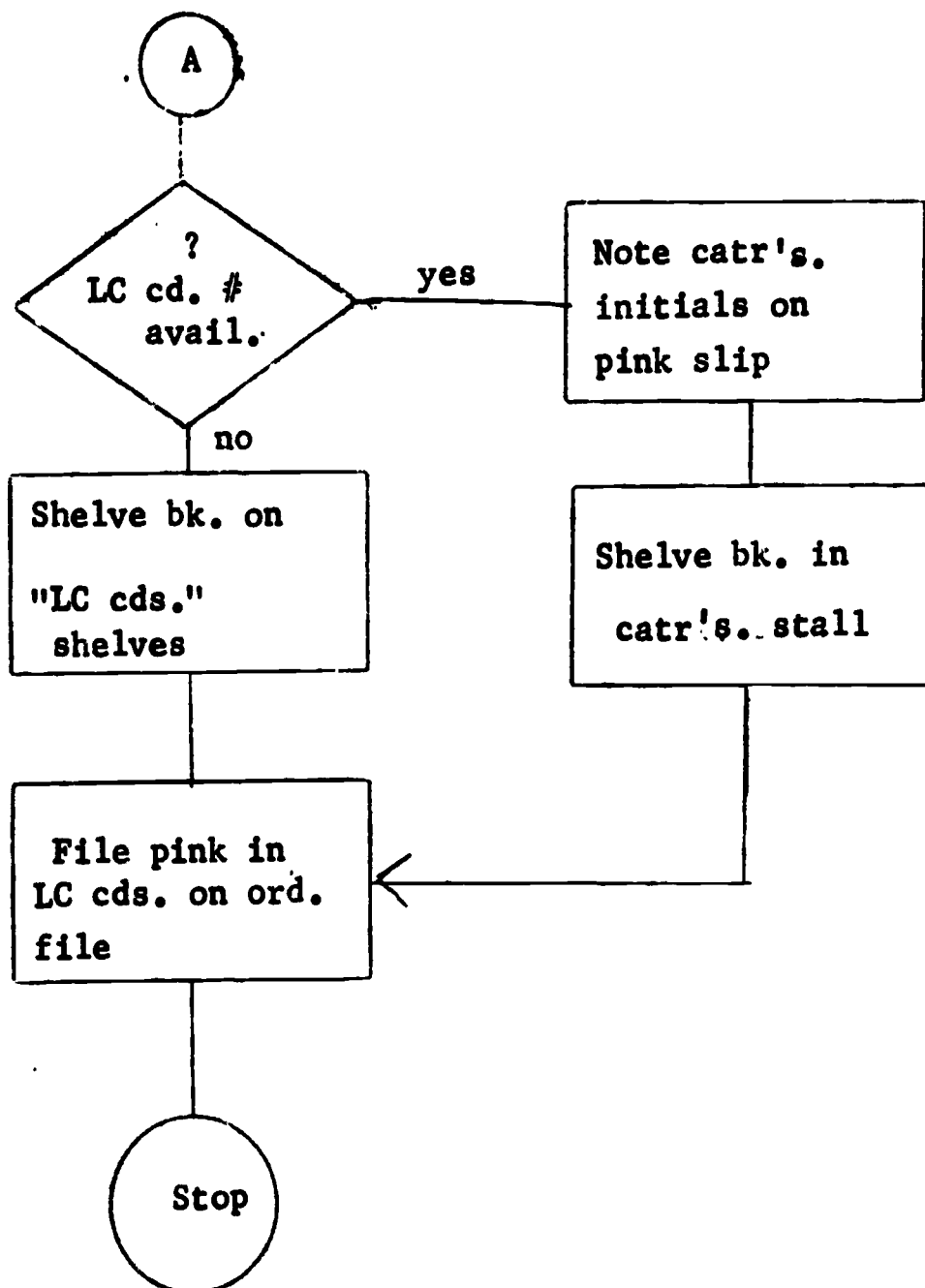
Library B--Library
Administration Office--
Receiving



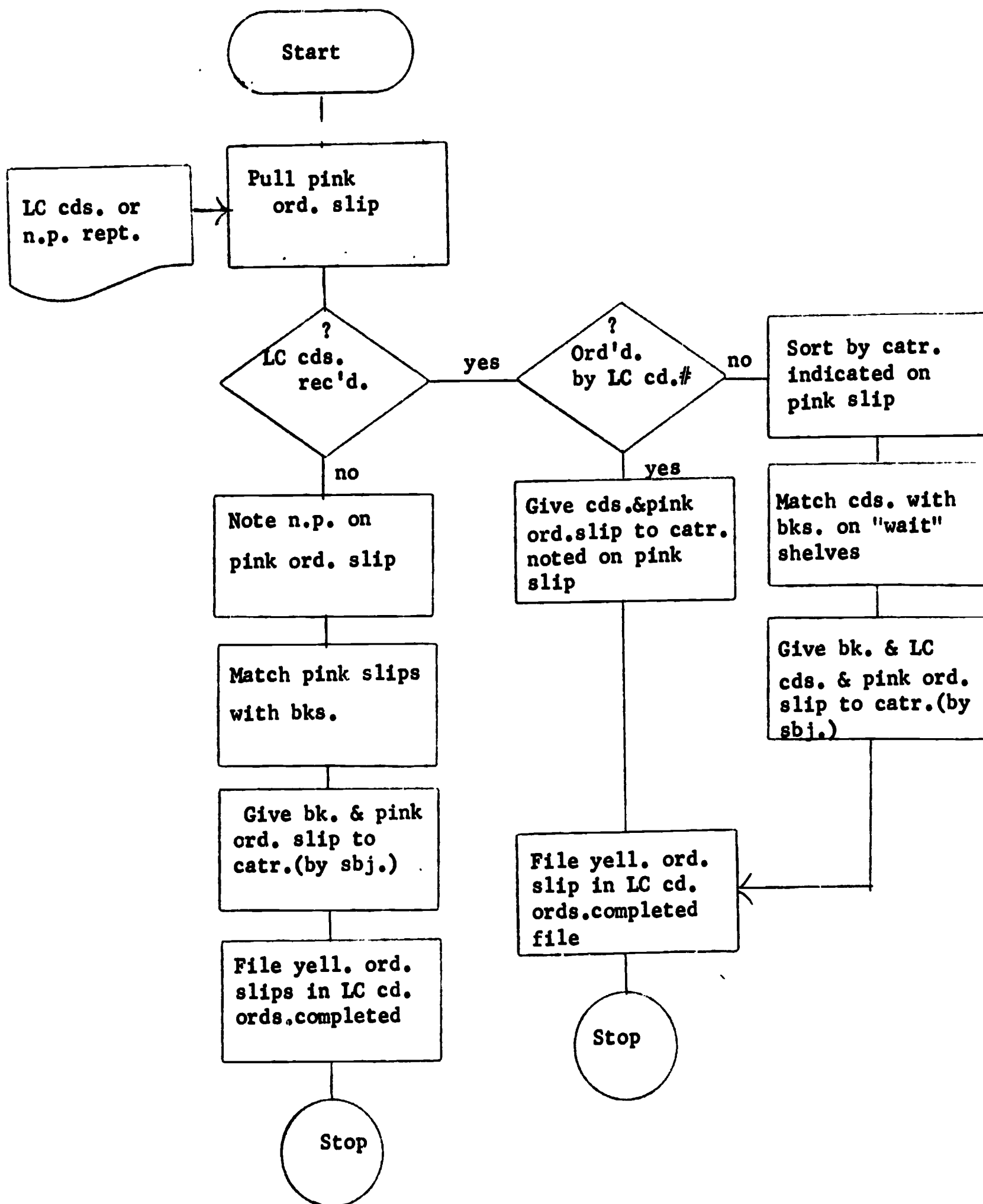
A - 18
Library L--Catalog Department--Ordering LC Cards



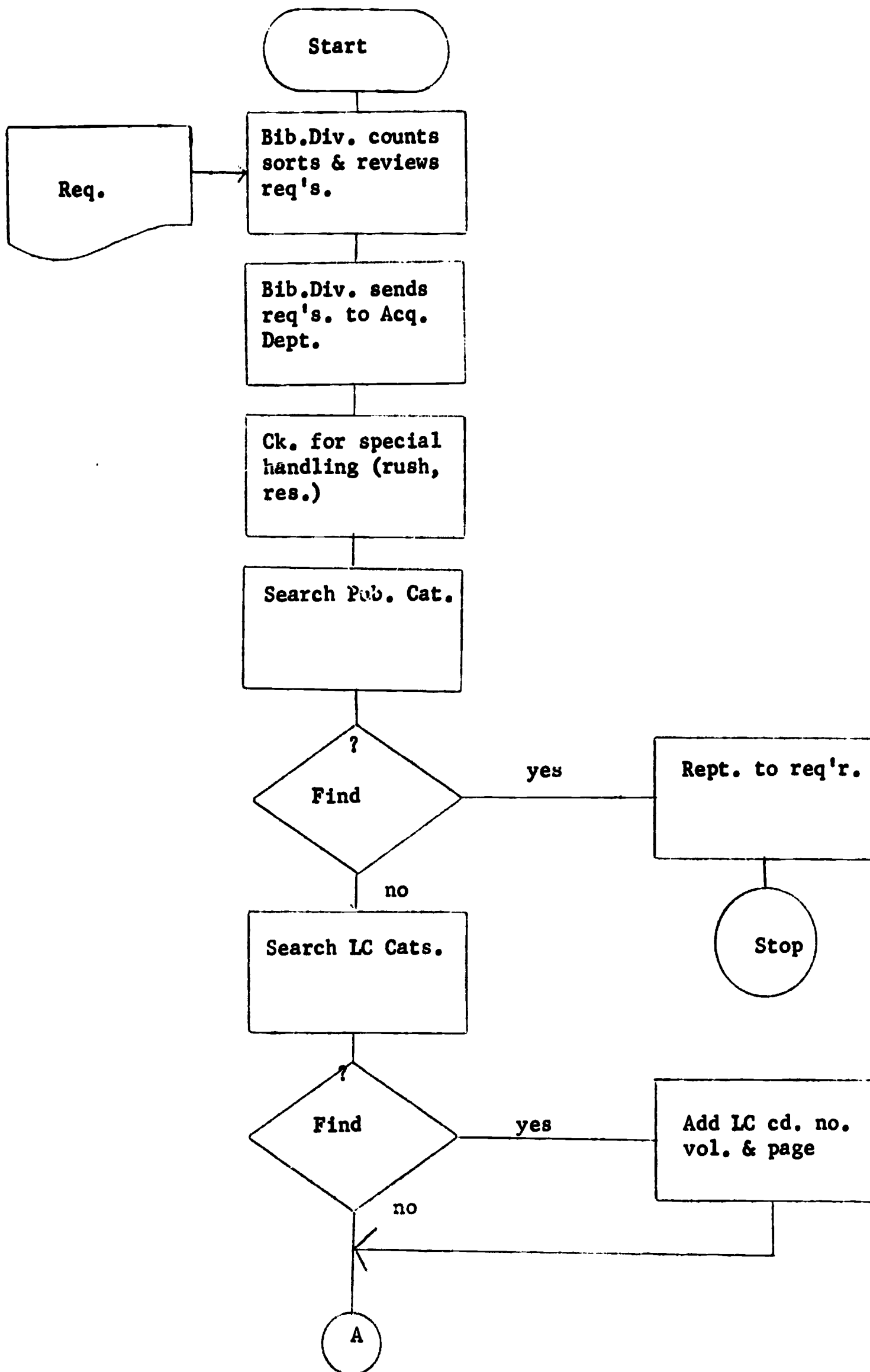
Library B--Catalog Department--Ordering LC Cards

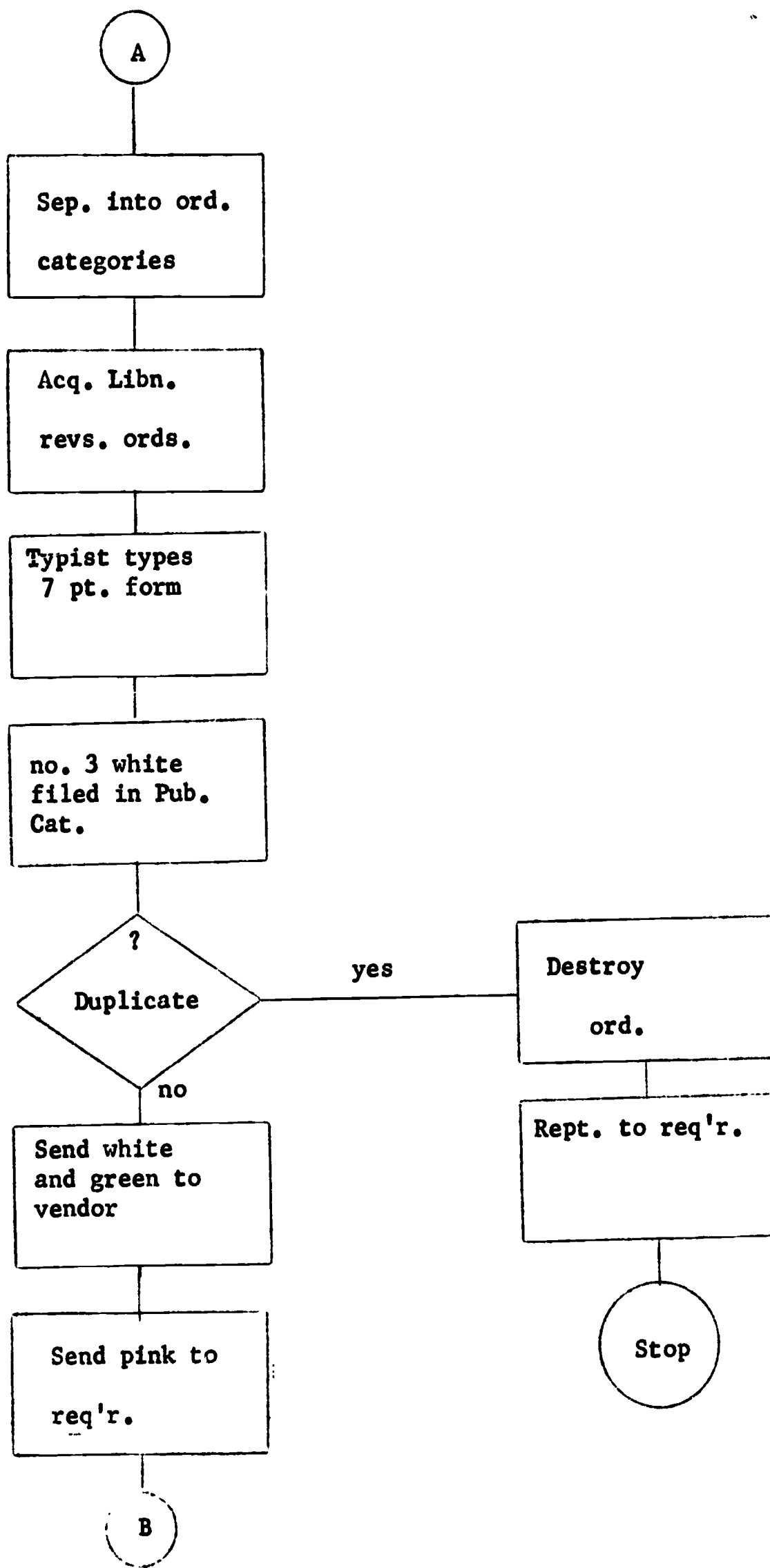


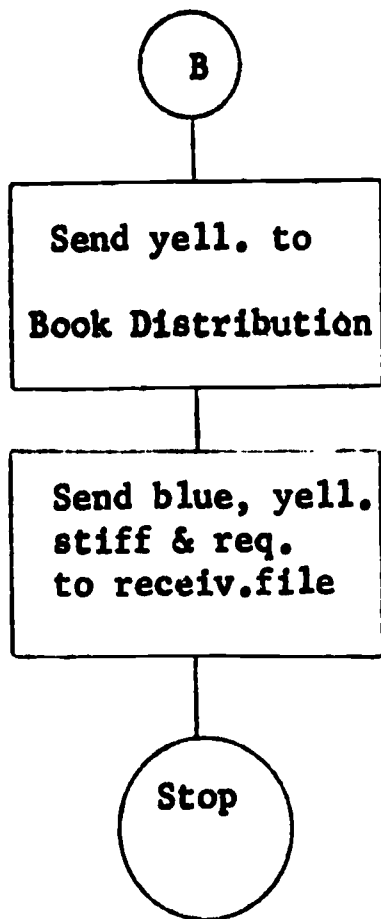
Library B--Catalog Department--Receiving LC Cards

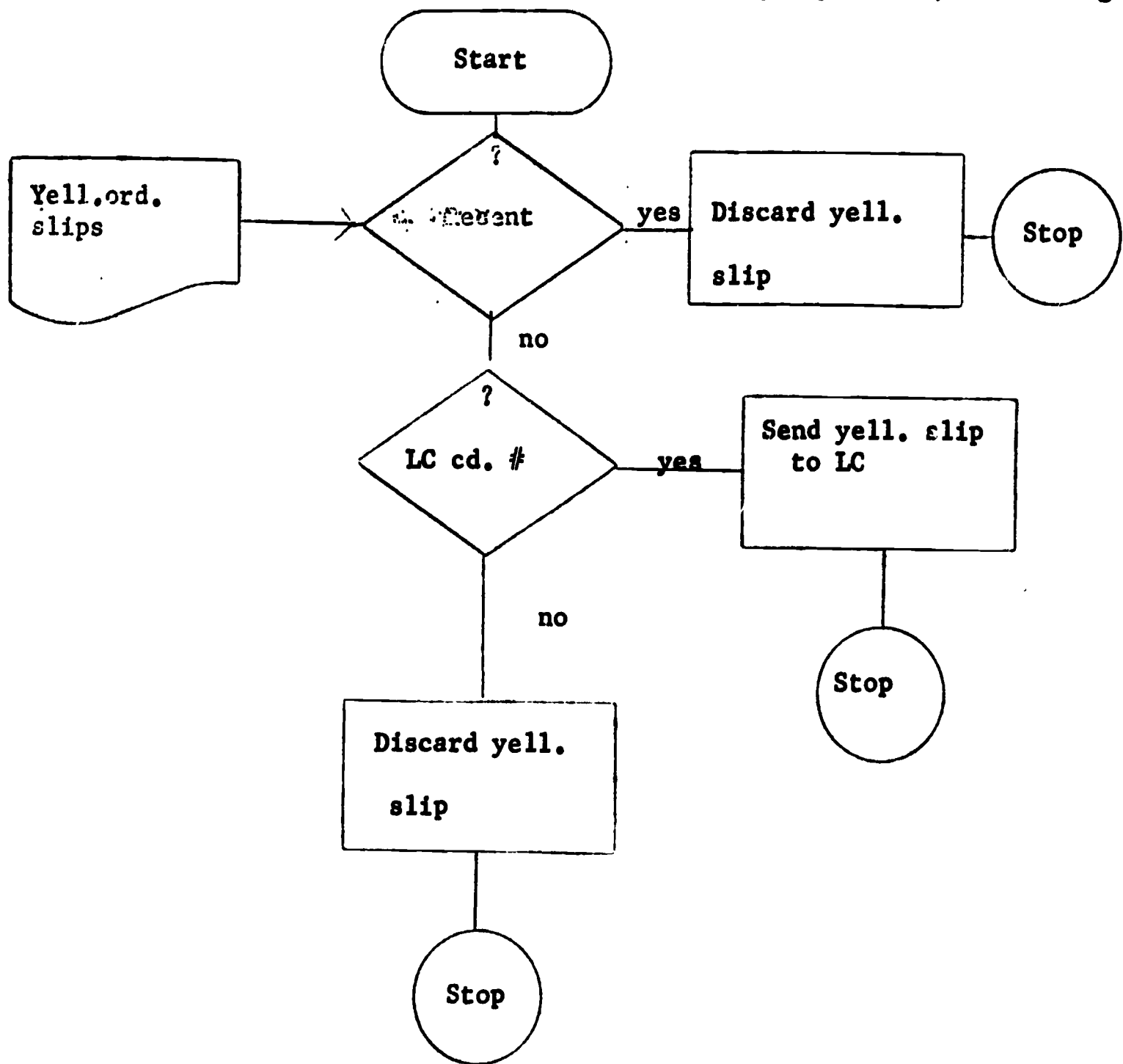


Library C--Acquisitions Department--Ordering

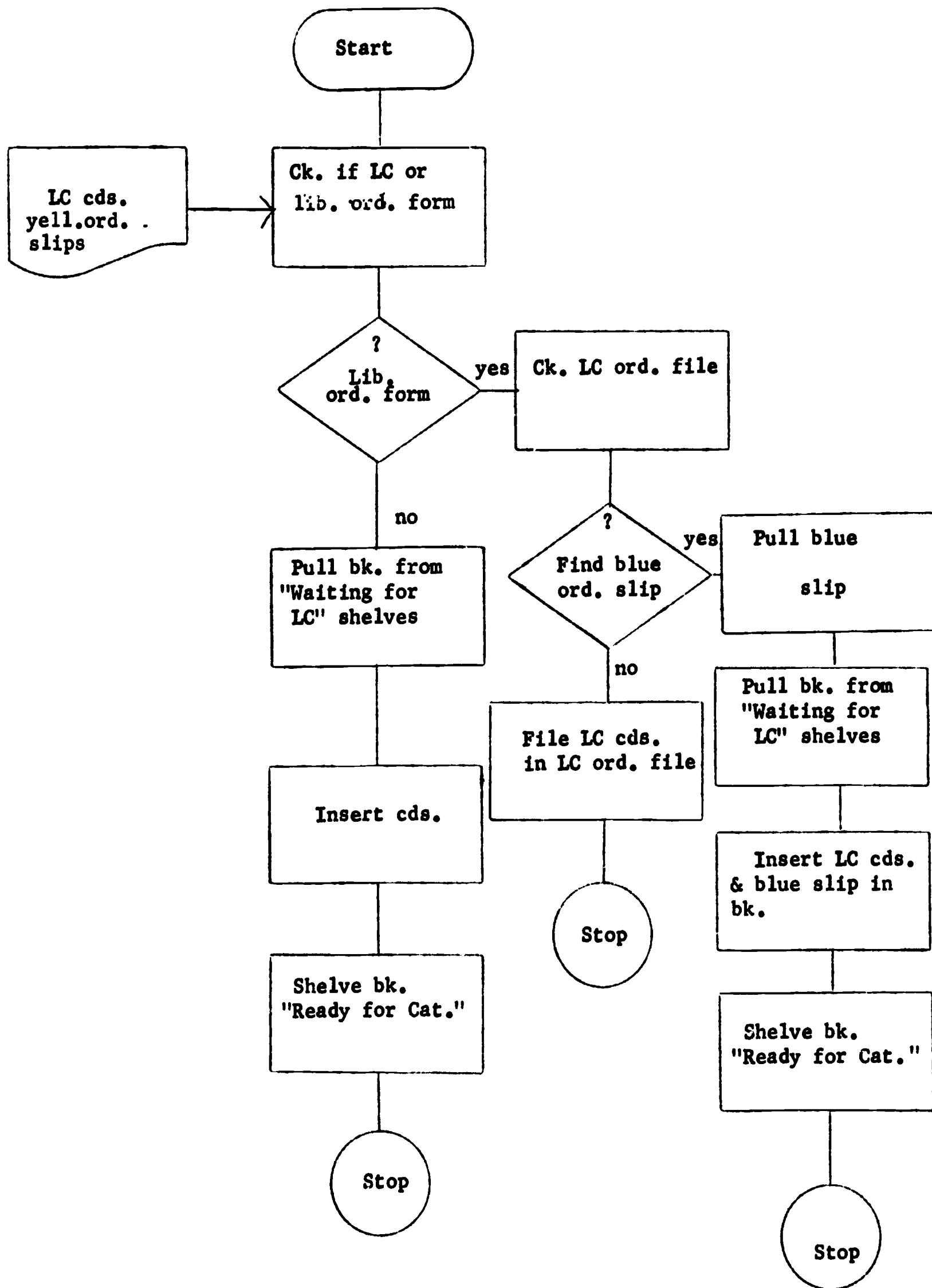




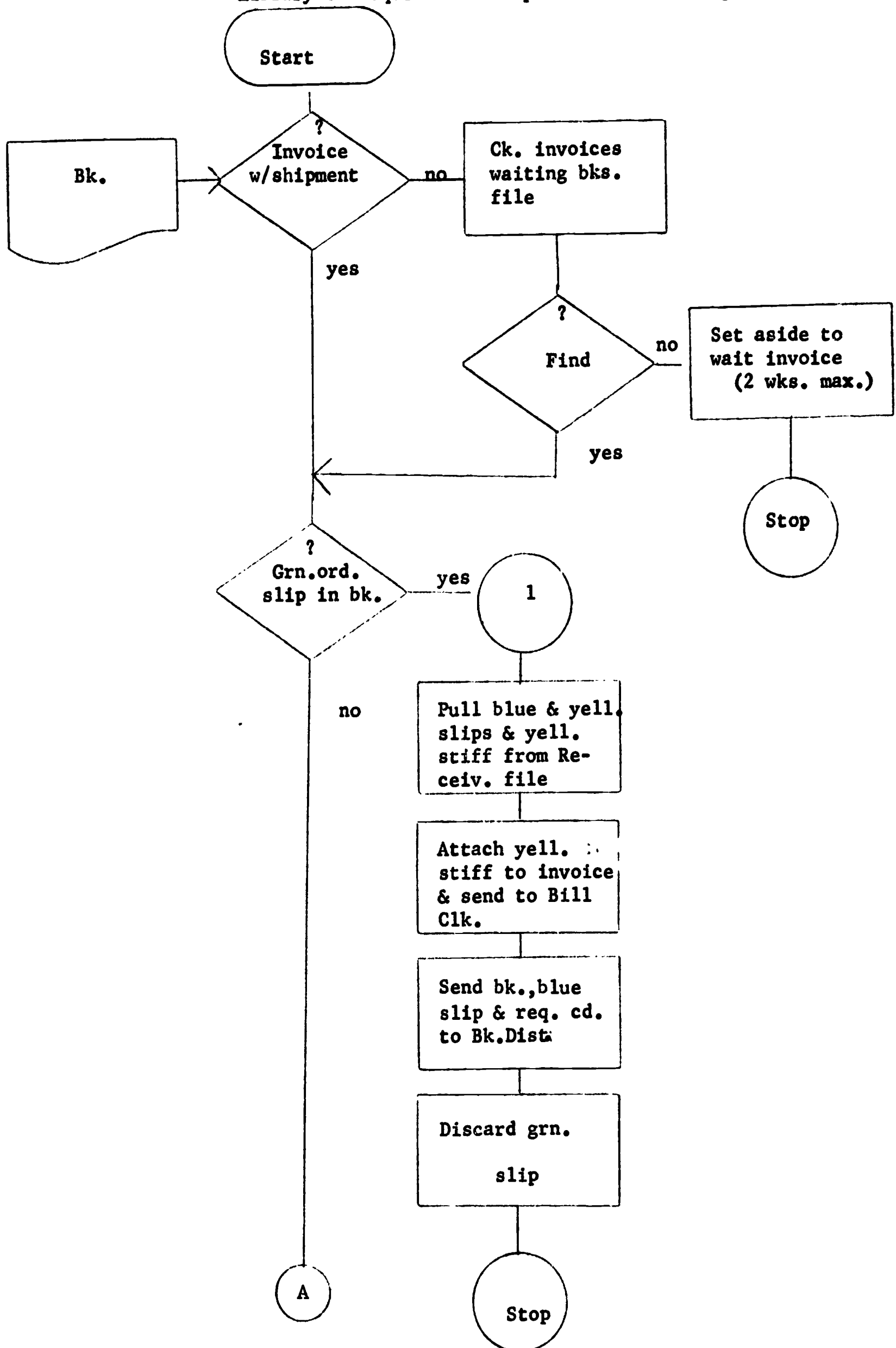


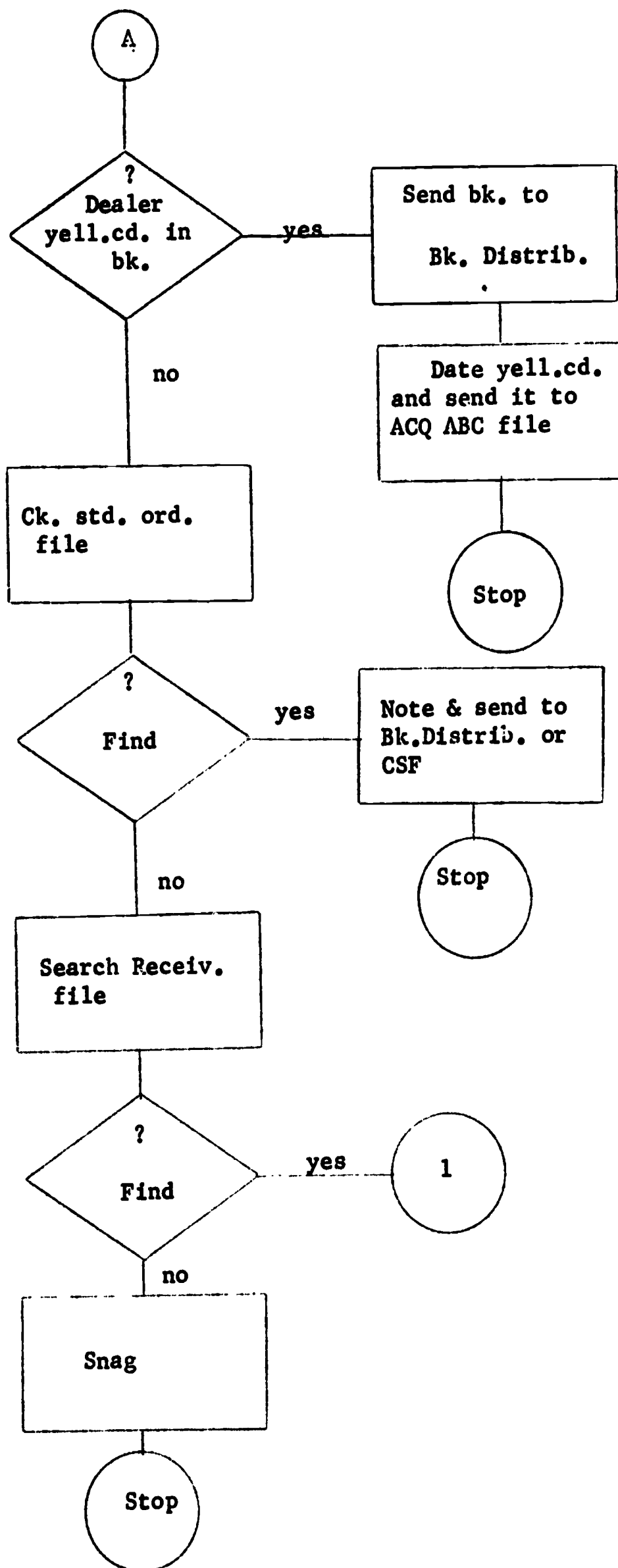


Library C--Book Distribution(Catalog Department)--Receiving

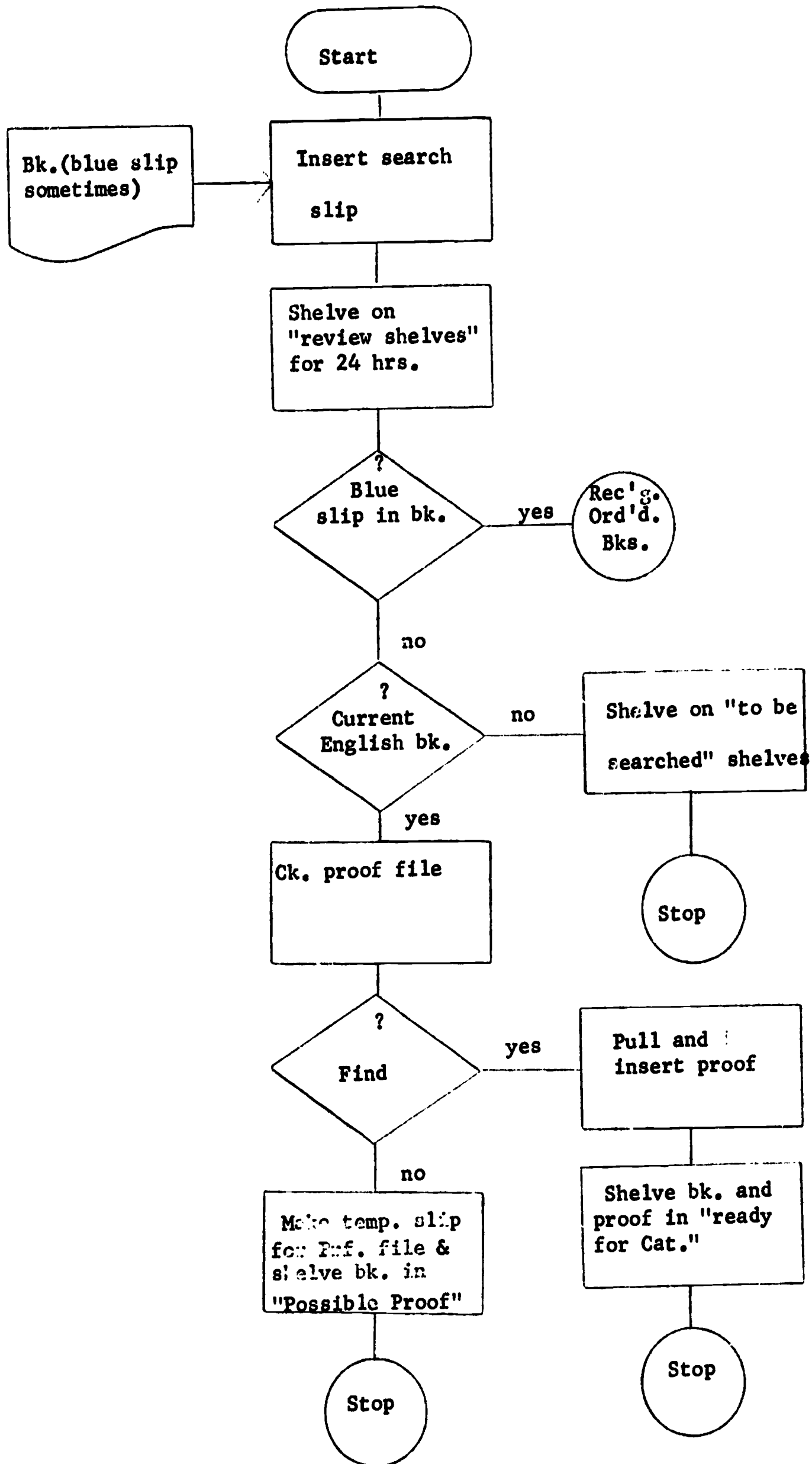


Library C--Acquisitions Department--Receiving

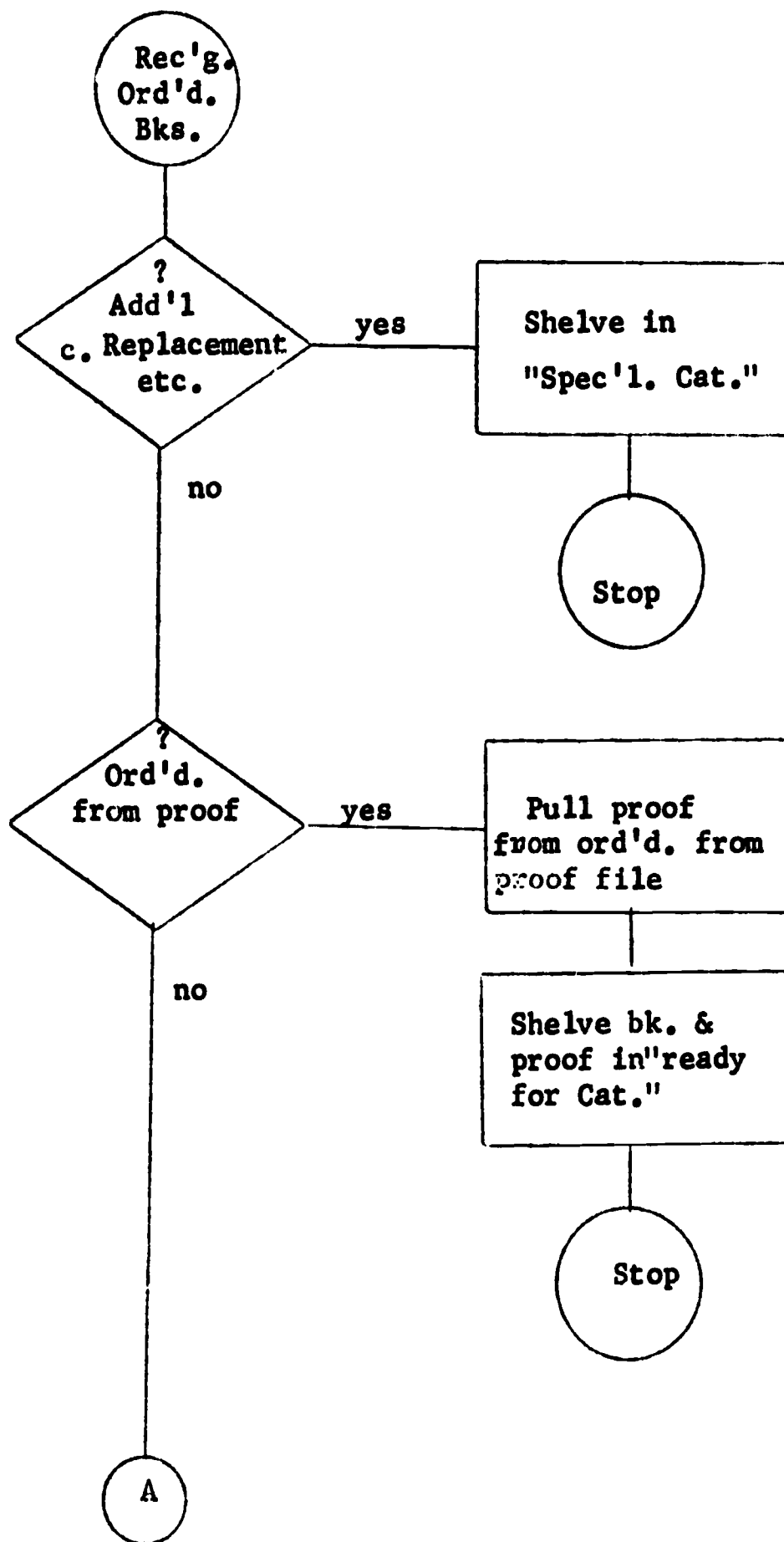


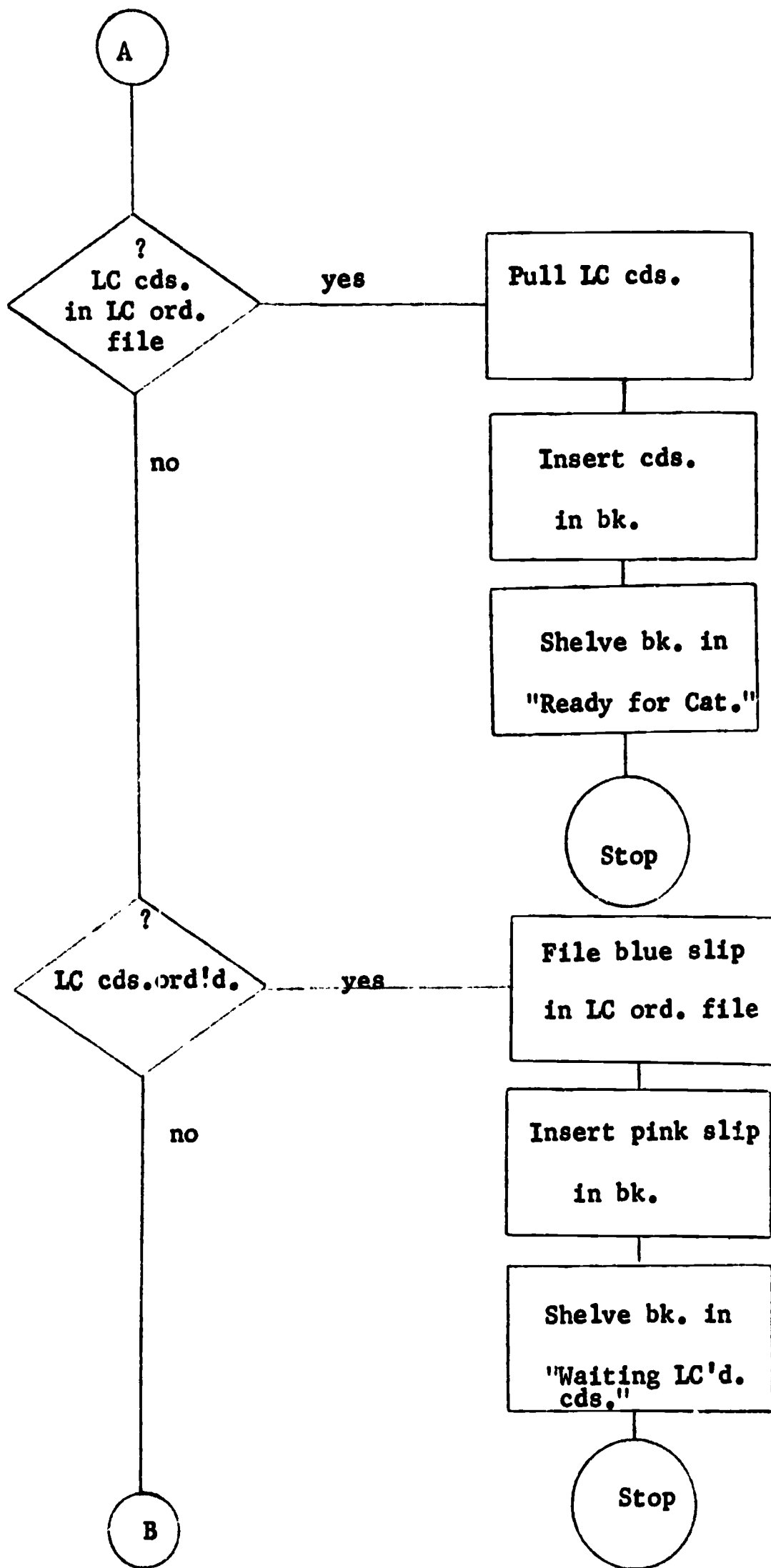


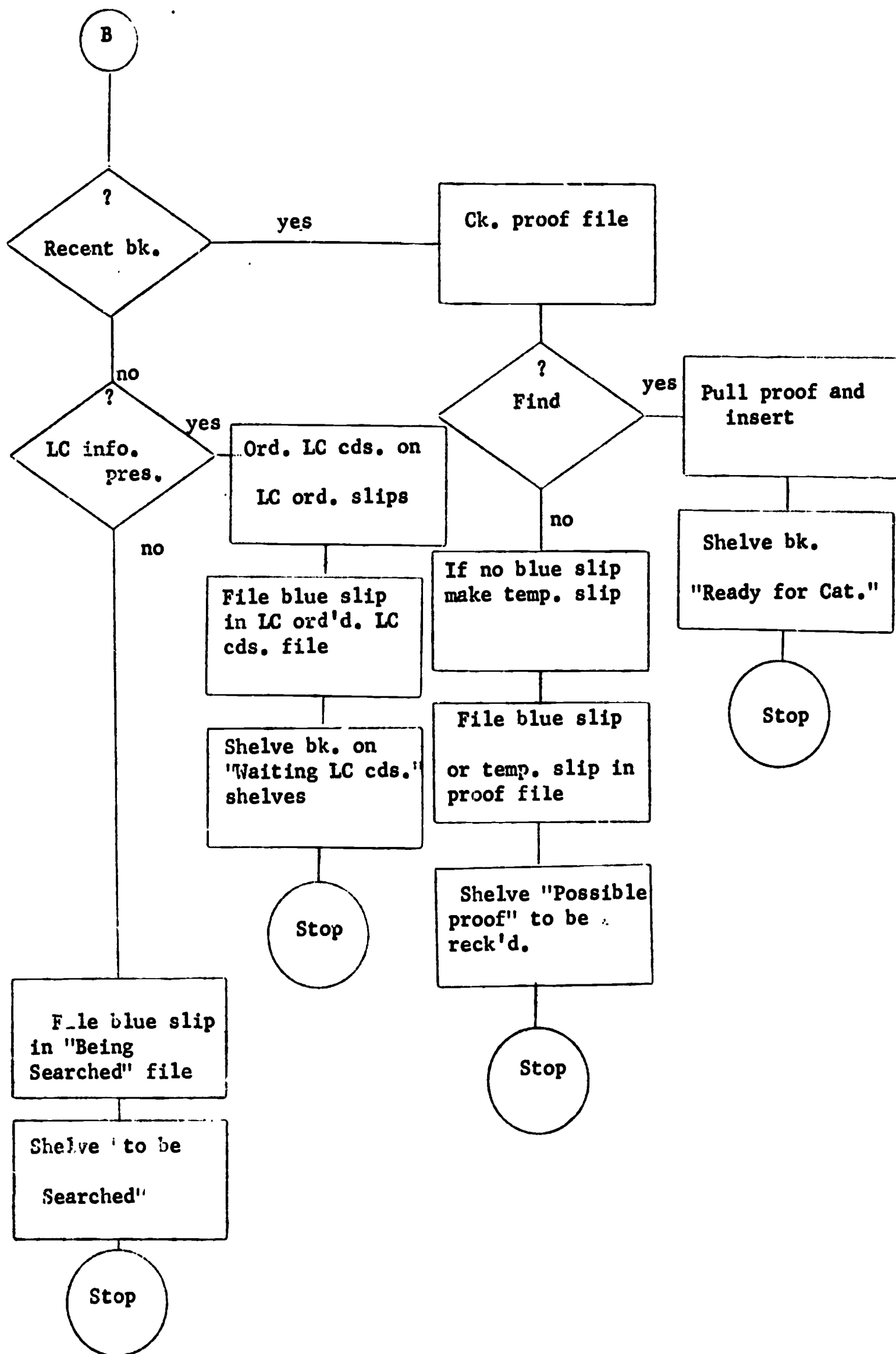
Library C--Book Distribution (Catalog Department)--Receiving



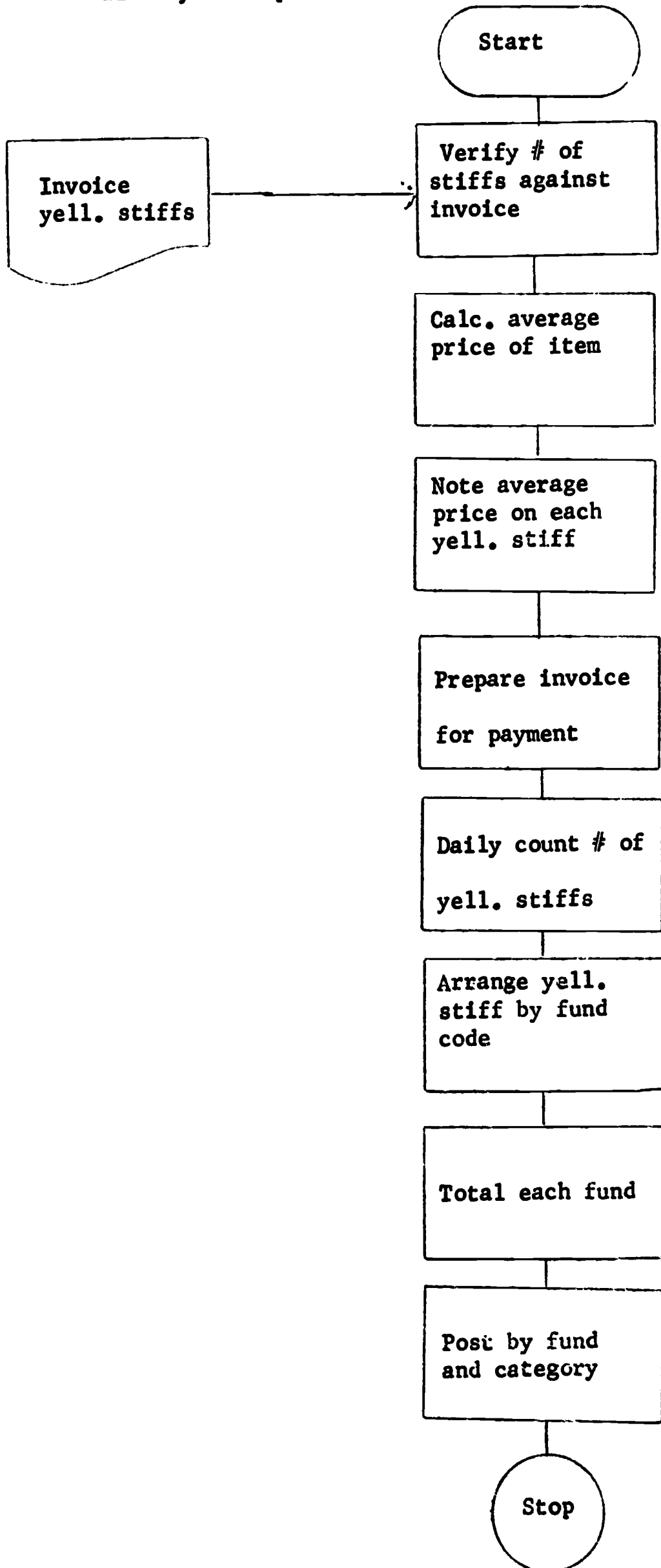
Library C--Book Distribution(Catalog Department)--Receiving



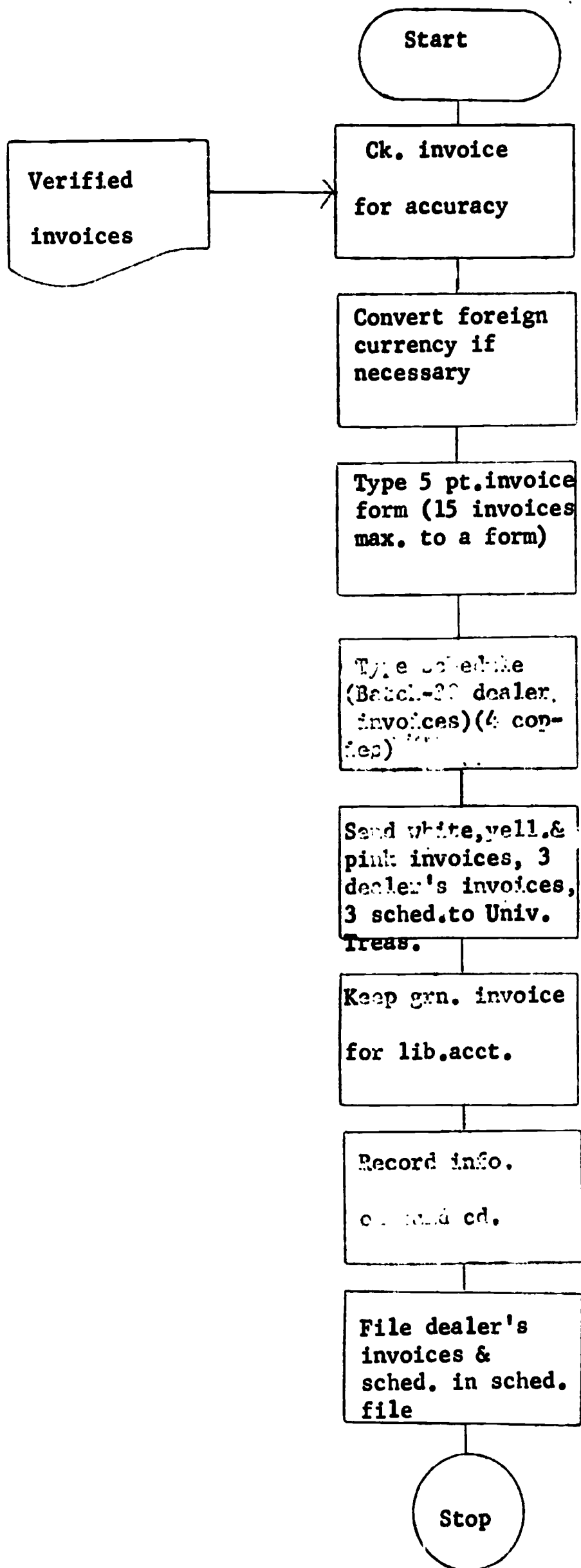


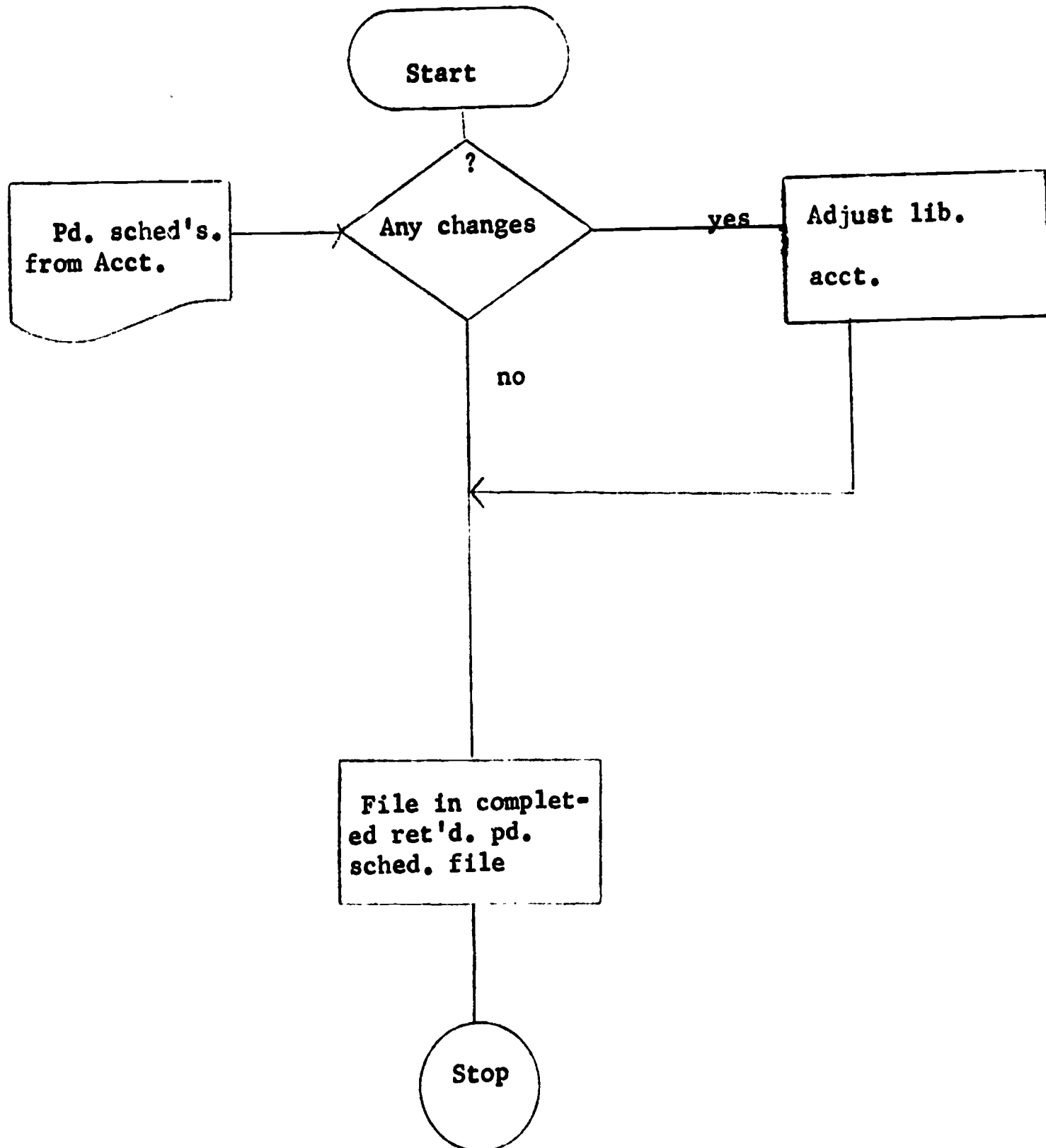


Library C--Acquisitions Department--Receiving

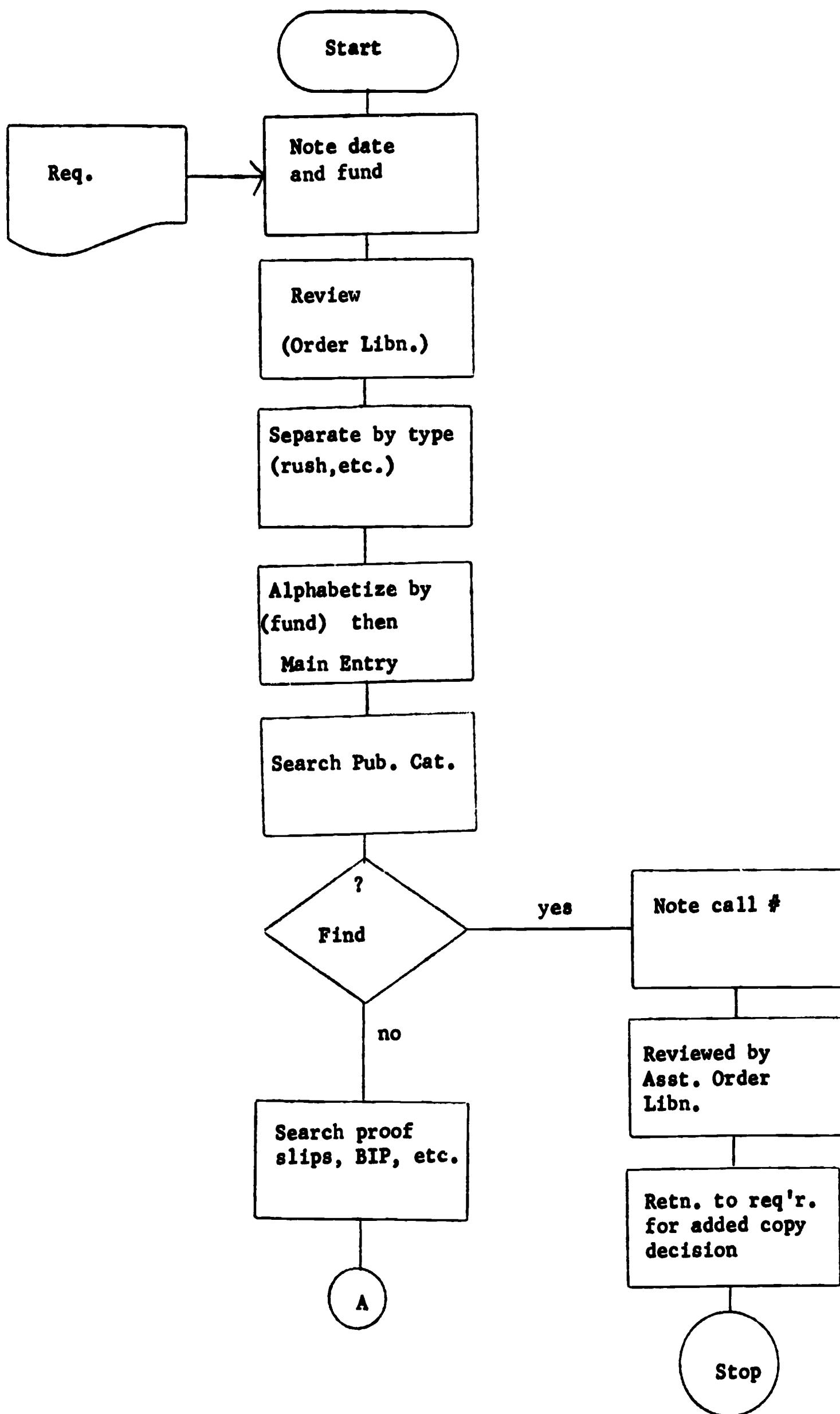


Library C--Acquisitions Department--Accounting

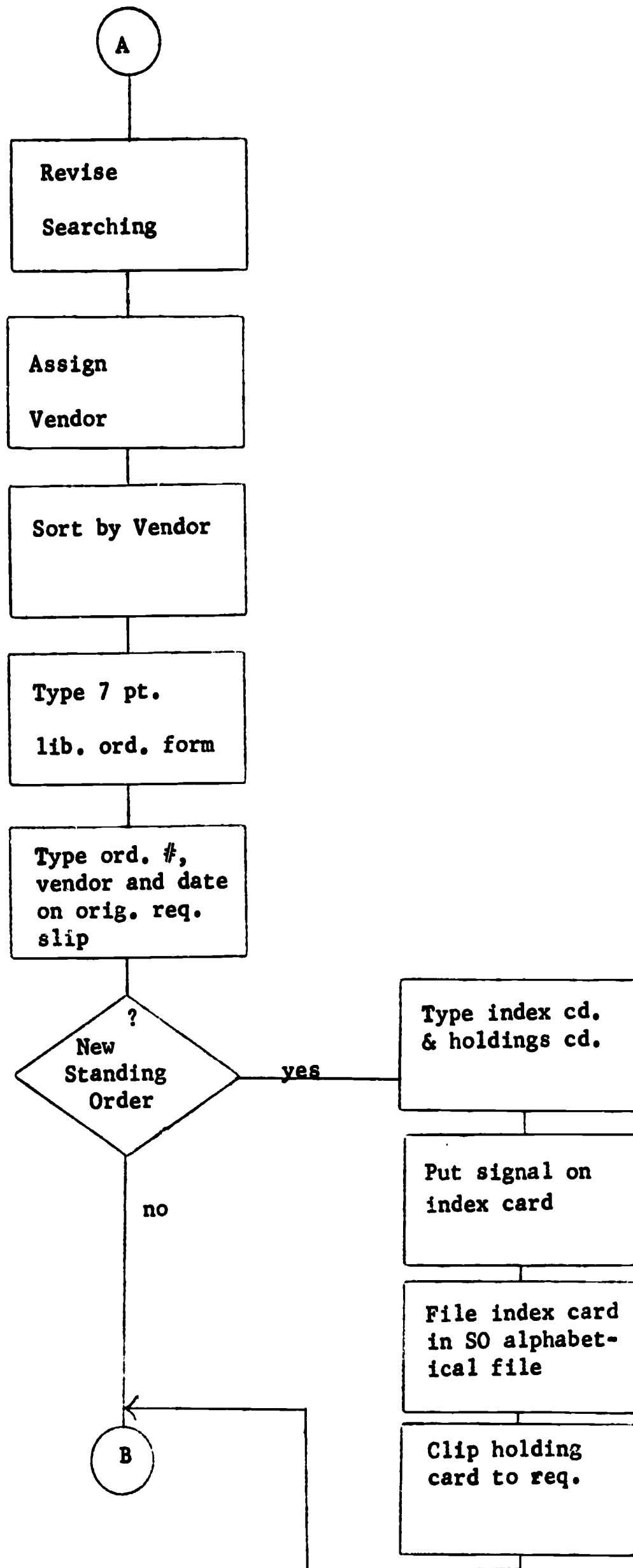




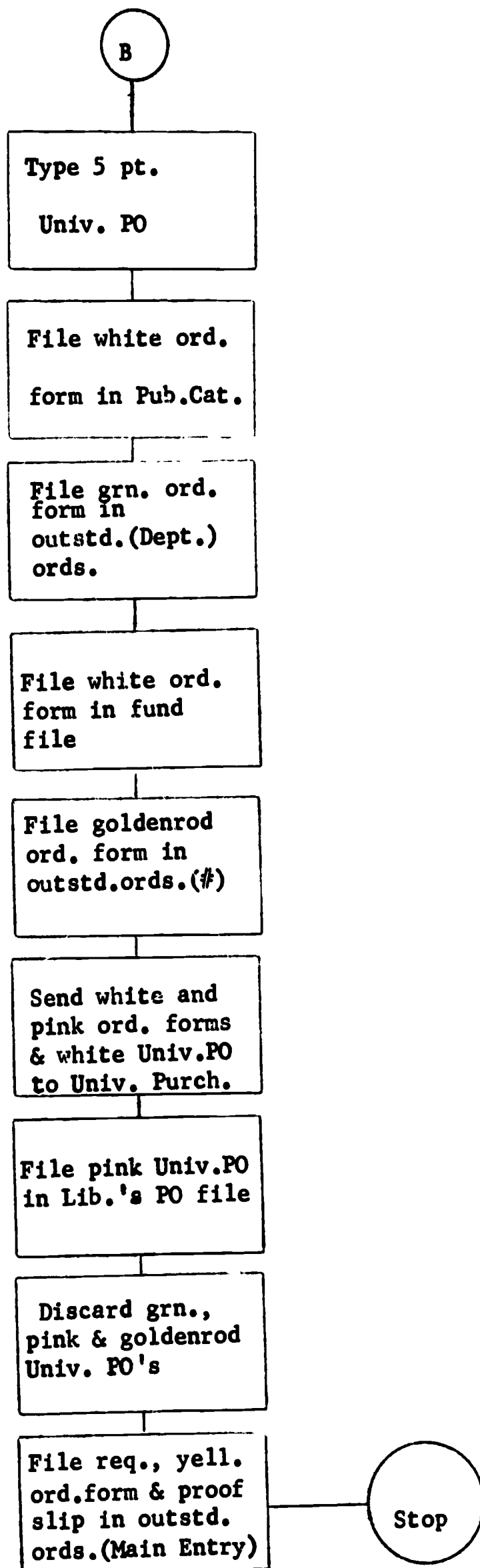
Library D--Acquisitions Department--Ordering

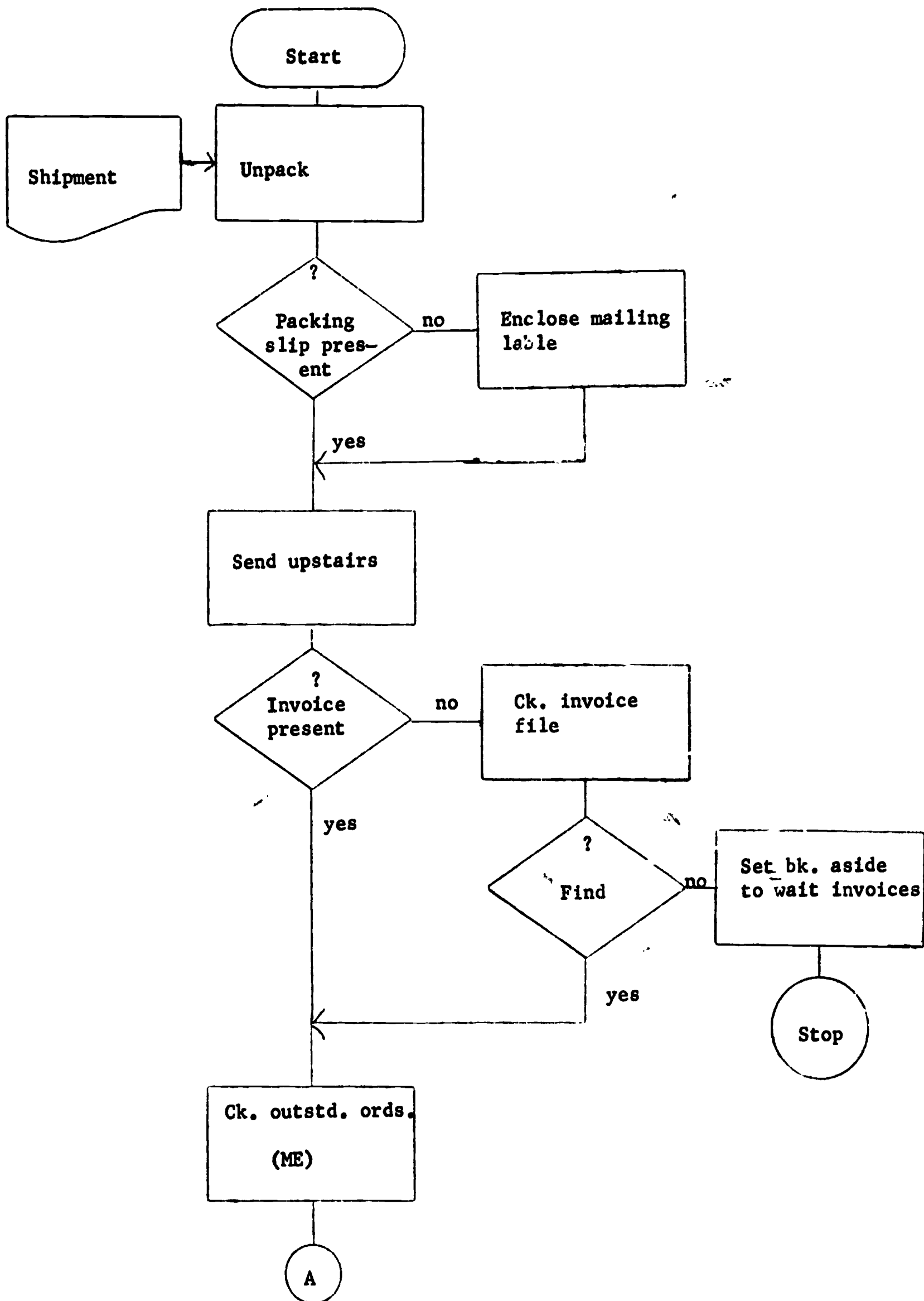


Library D--Acquisitions Department--Ordering

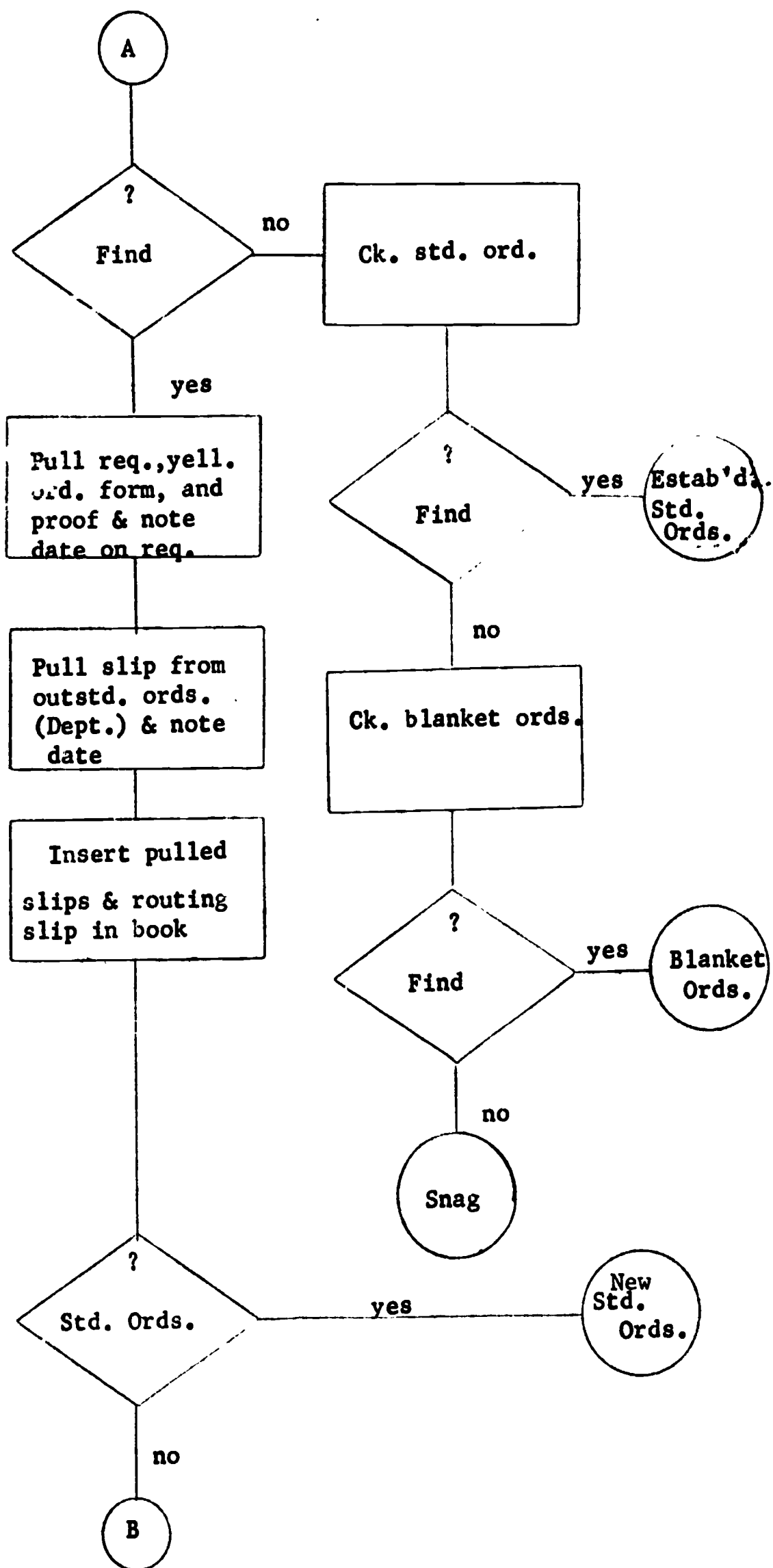


Library D--Acquisitions Department--Ordering

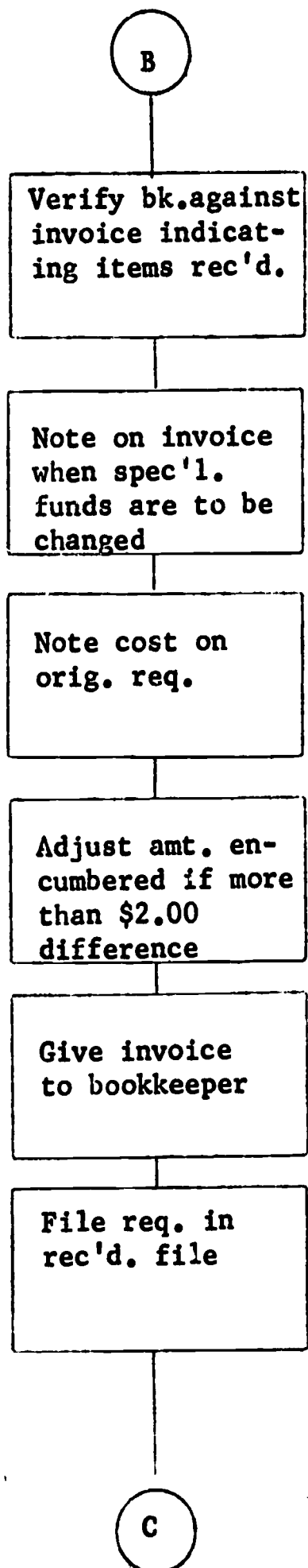




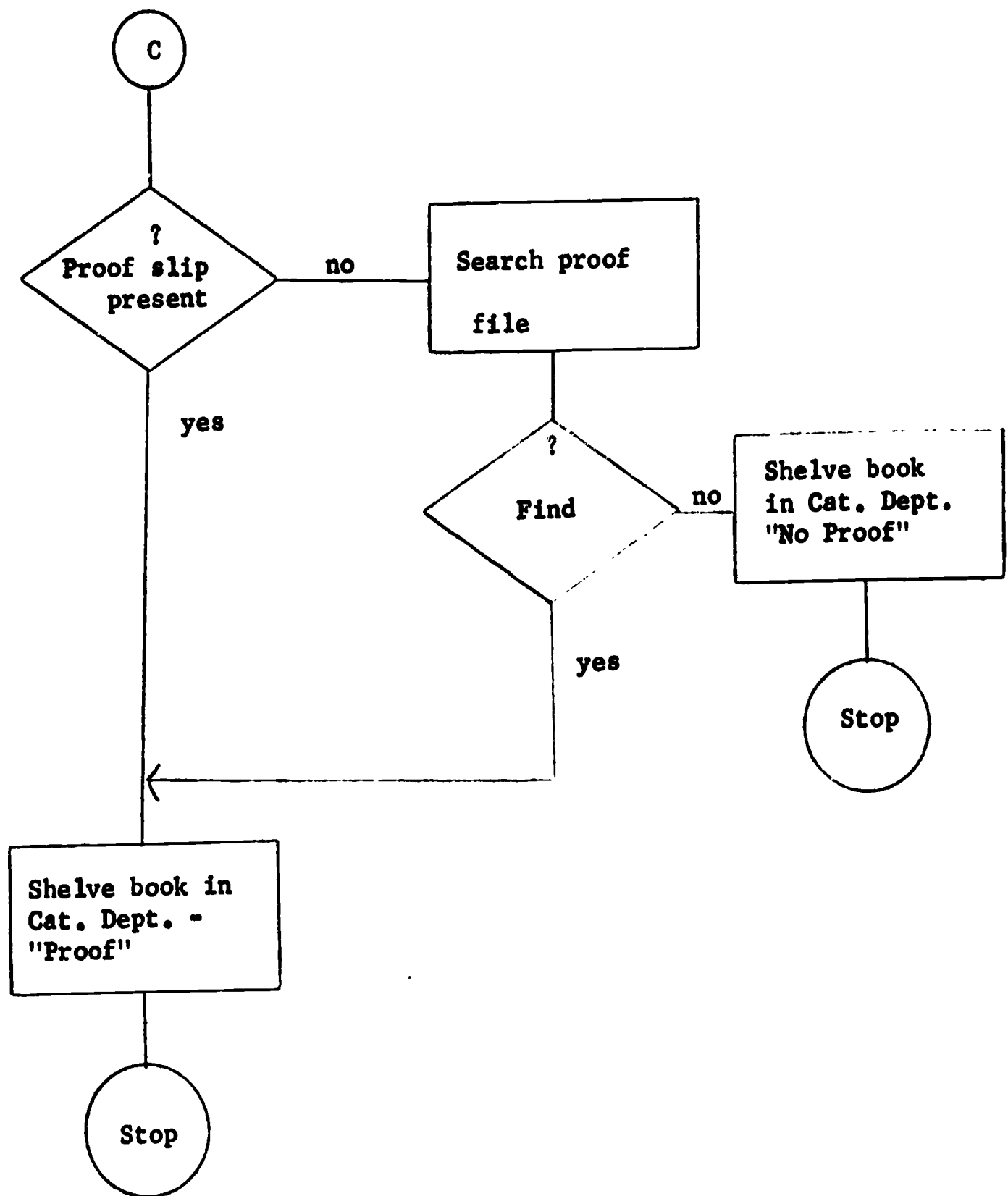
Library D--Acquisitions Department--Receiving



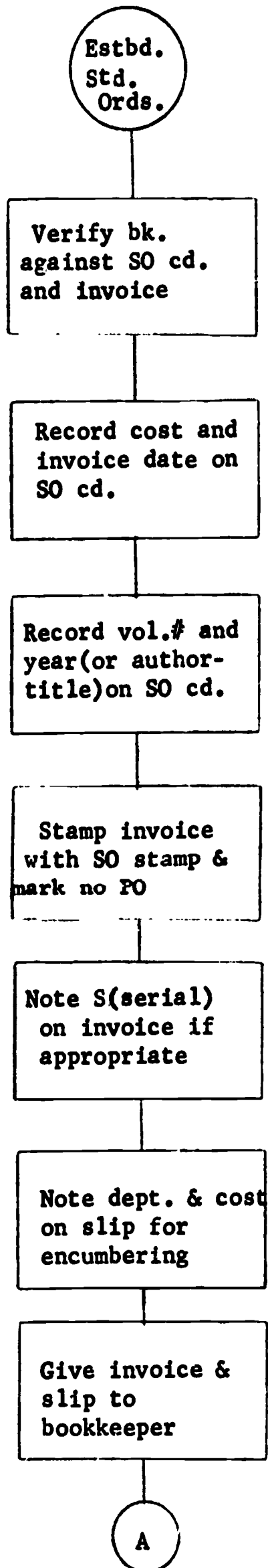
Library D--Acquisitions Department--Receiving

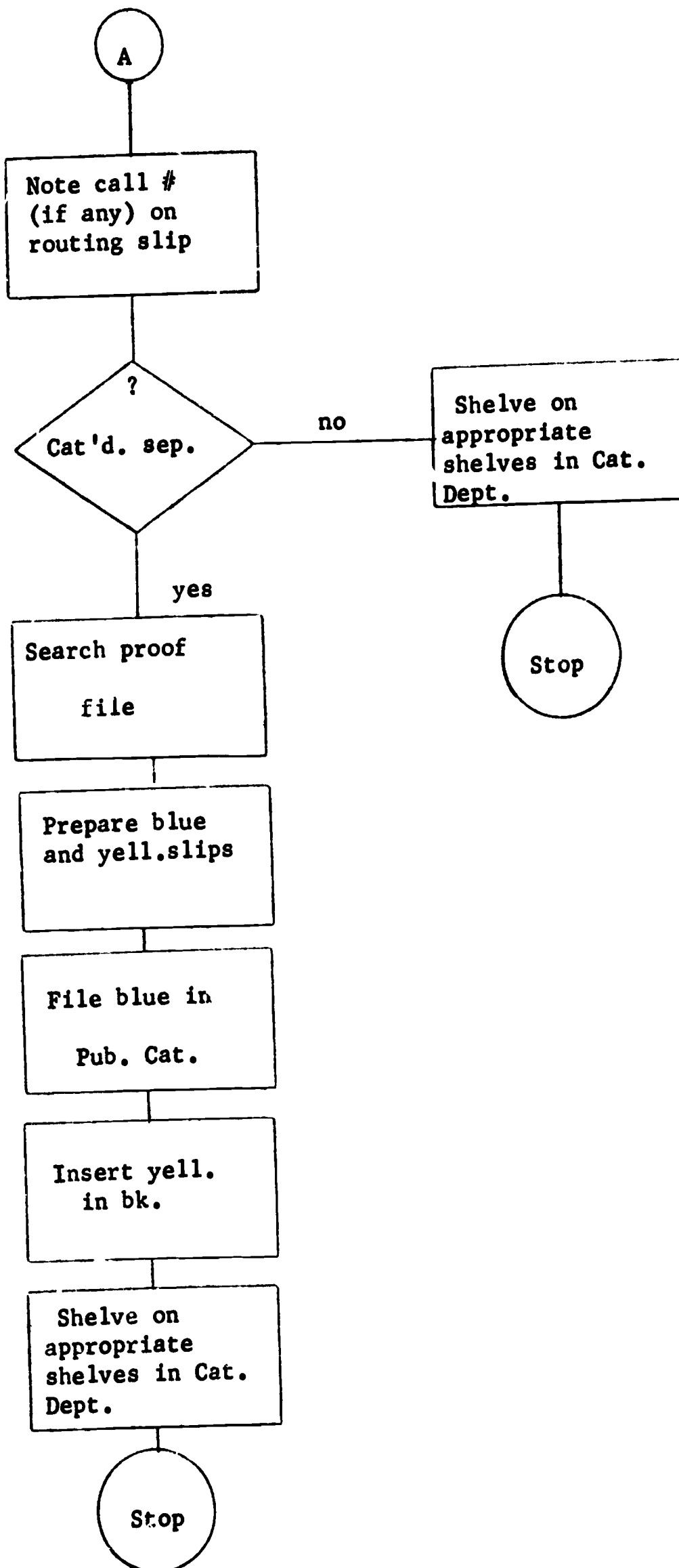


Library D--Acquisitions Department--Receiving

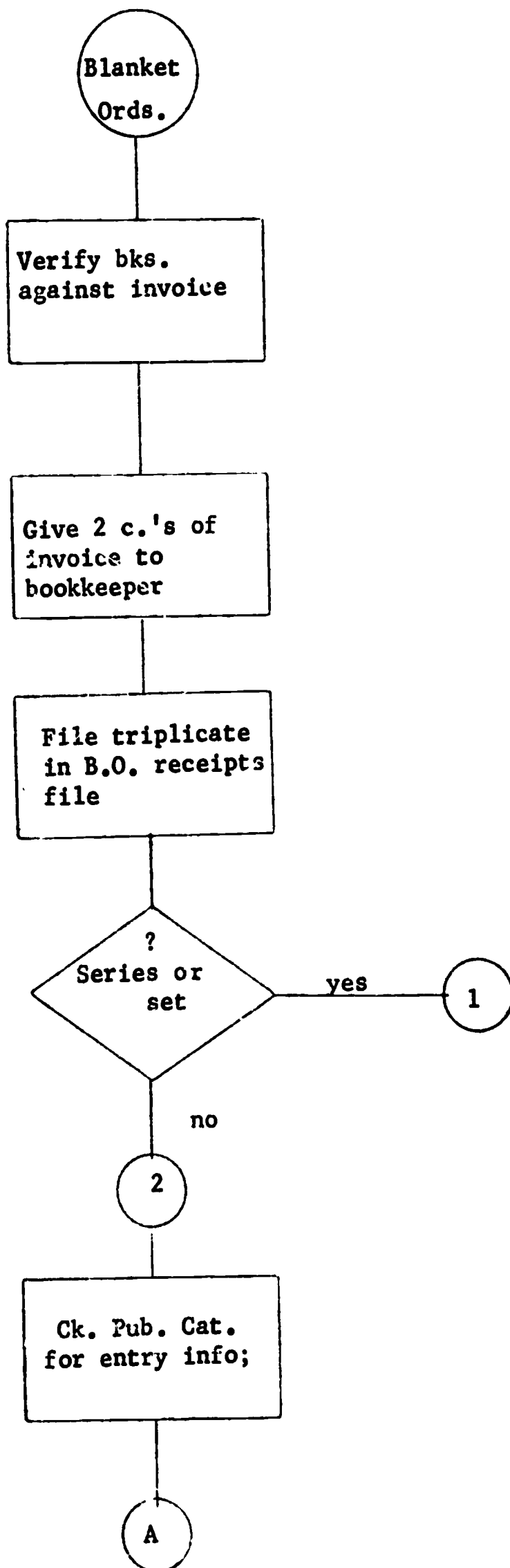


Library D--Acquisitions Department--Receiving Established
Standing Orders

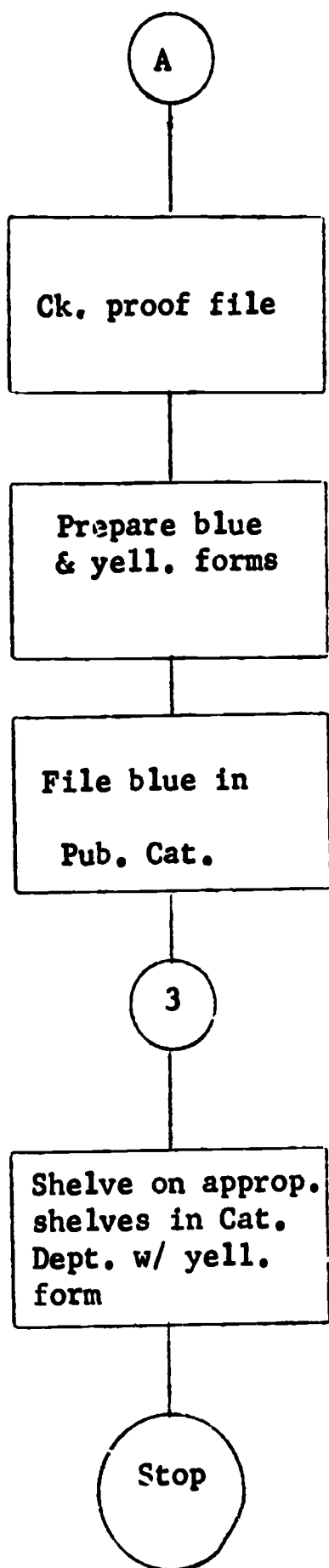


Library D--Acquisitions Department--Receiving Established
Standing Orders

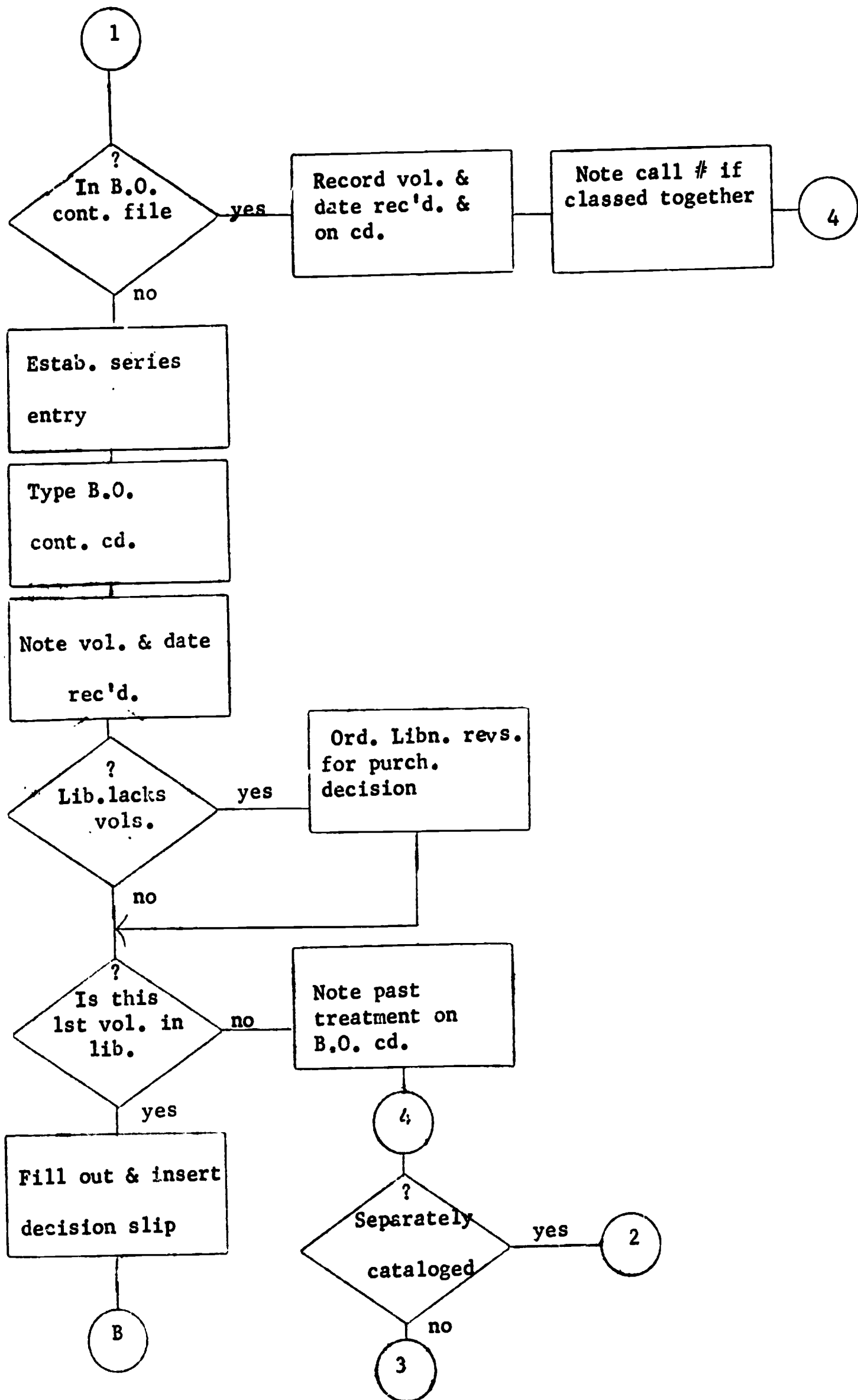
Library D--Acquisitions Department--Receiving Blanket Orders



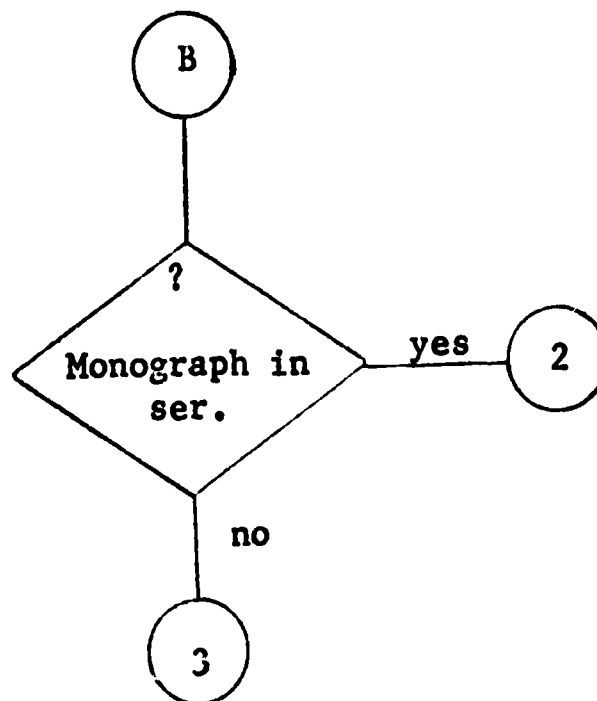
Library D--Acquisitions Department--Receiving Blanket Orders



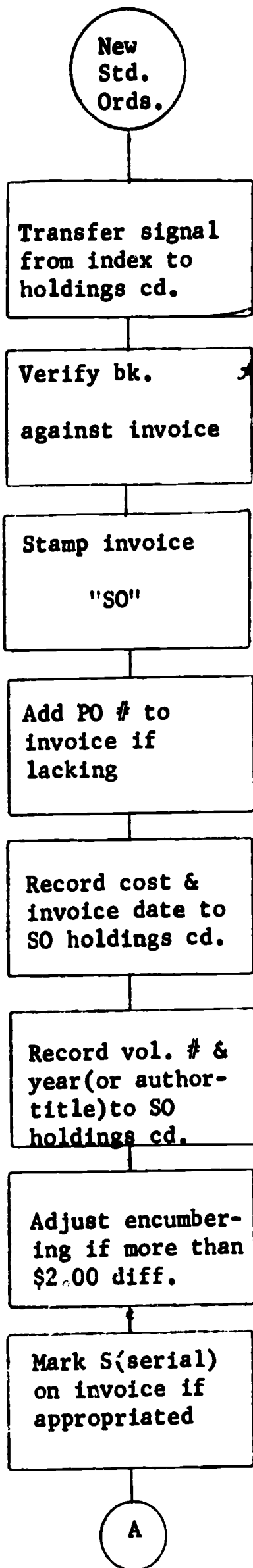
A - 45
Library D--Acquisitions Department--Receiving Blanket Orders



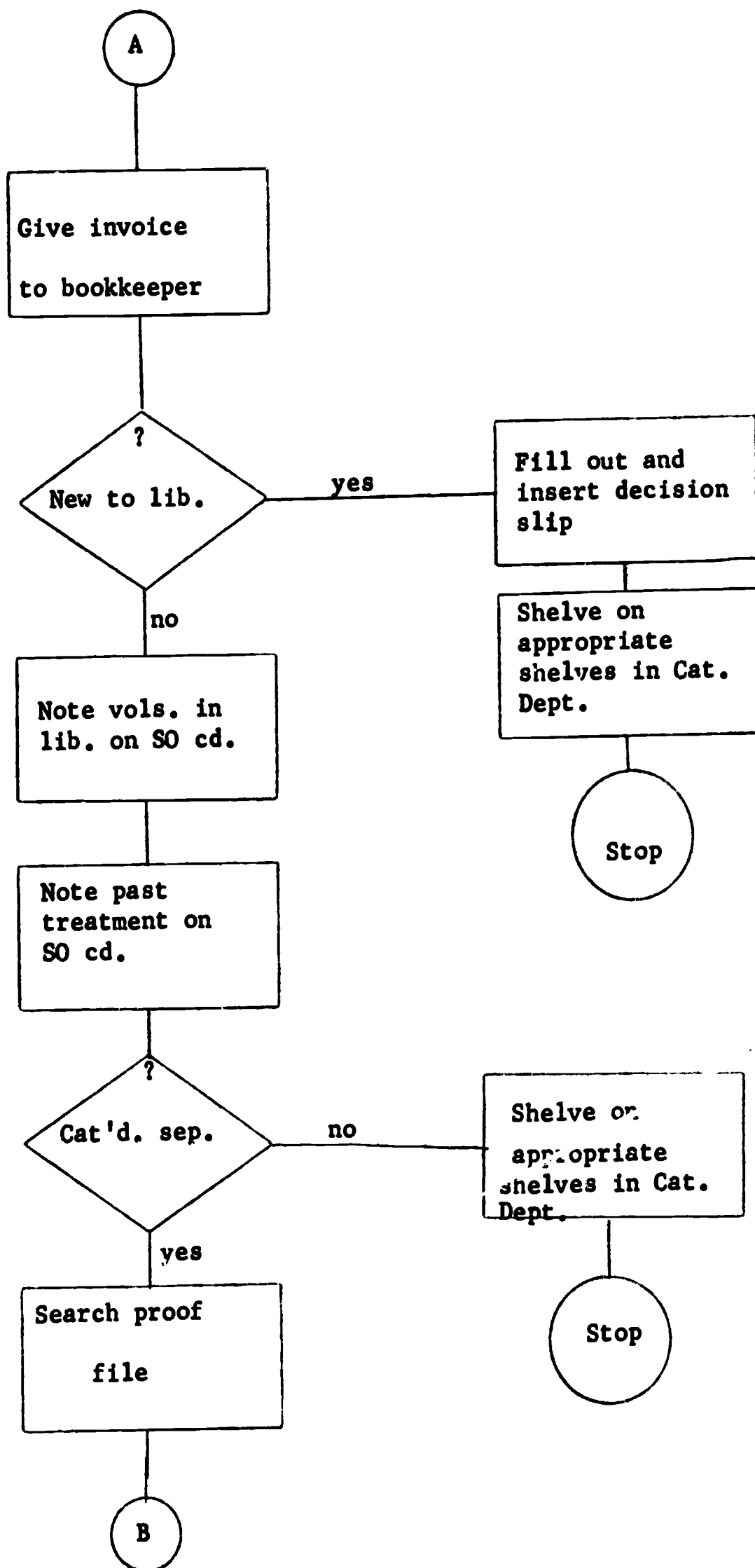
Library D--Acquisitions Department--Receiving Blanket Orders



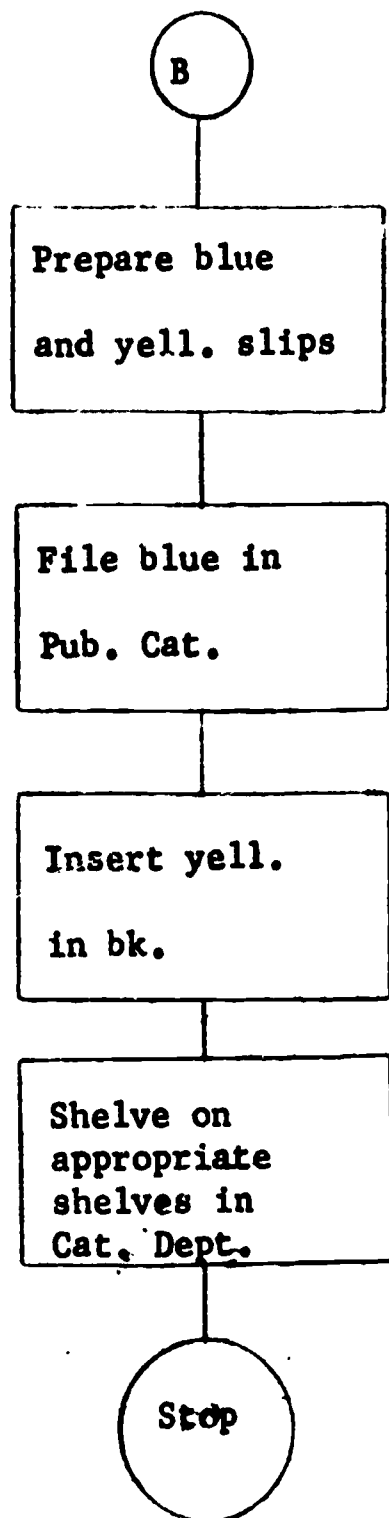
Library D--Acquisitions Department--Receiving New Standing Orders



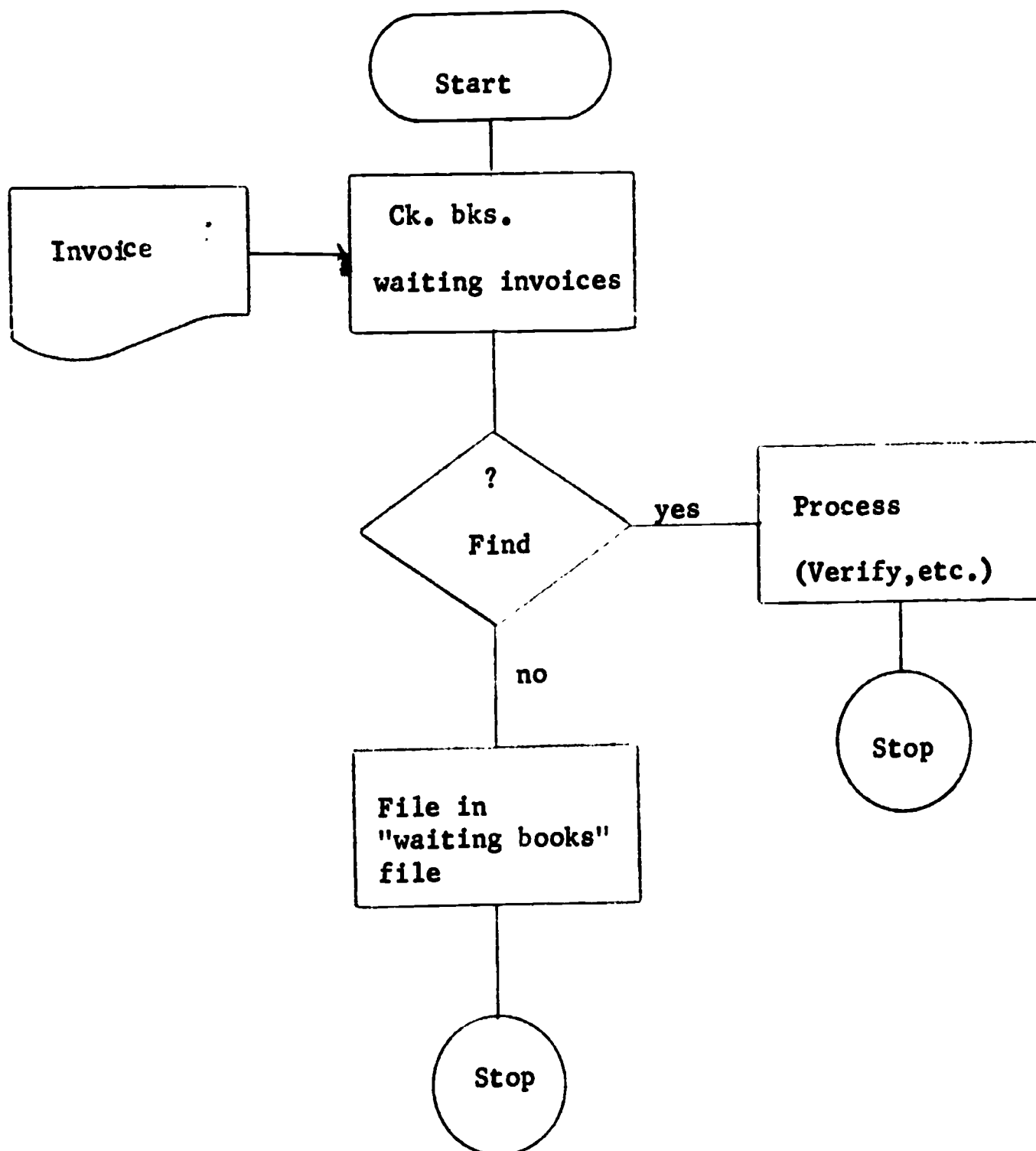
Library D--Acquisitions Department--Receiving New Standing Orders



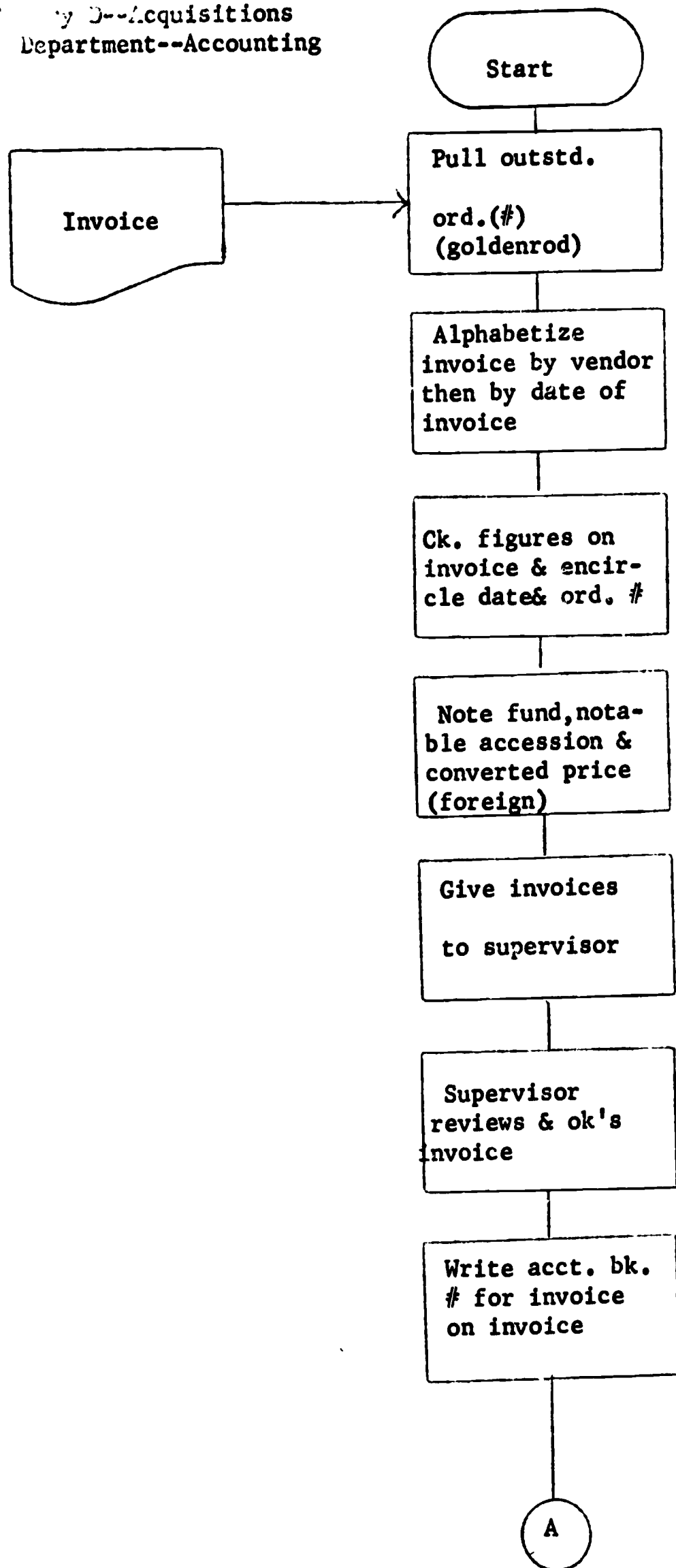
Library D--Acquisitions Department--Receiving New Standing Orders



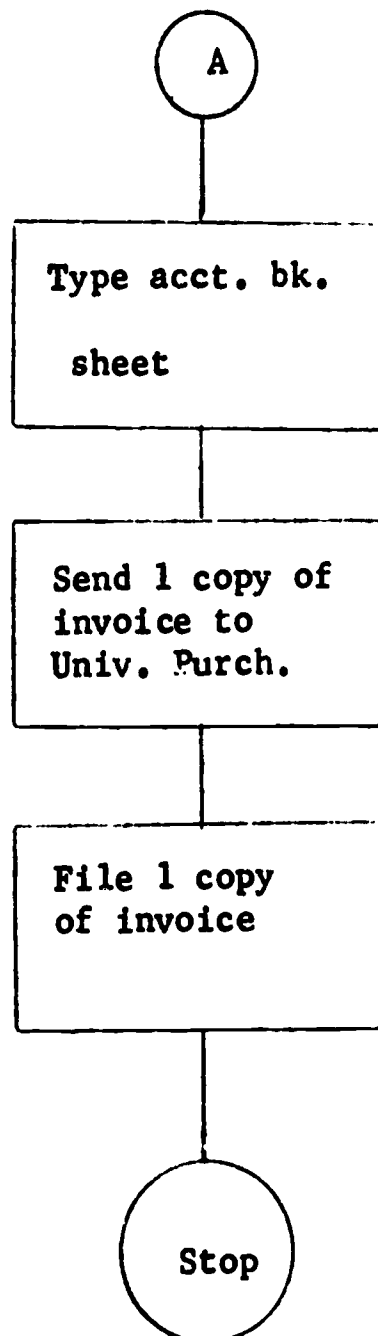
A - 51
Library D--Acquisitions Department--Invoice Receiving



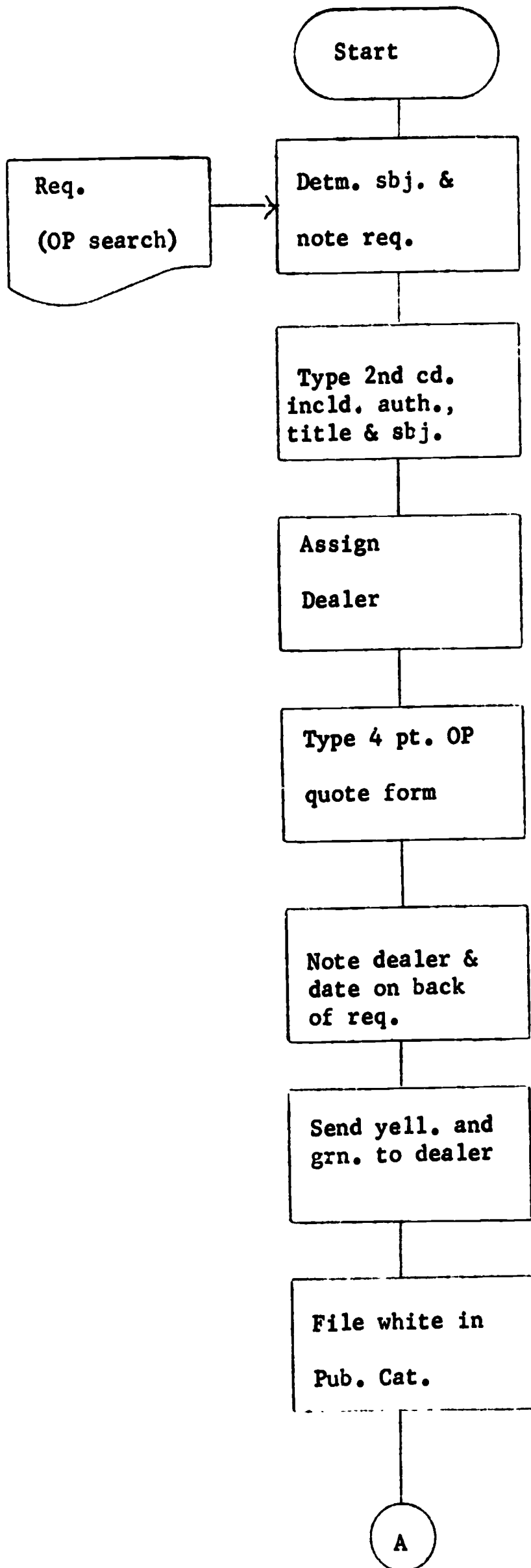
1. by D--Acquisitions
Department--Accounting



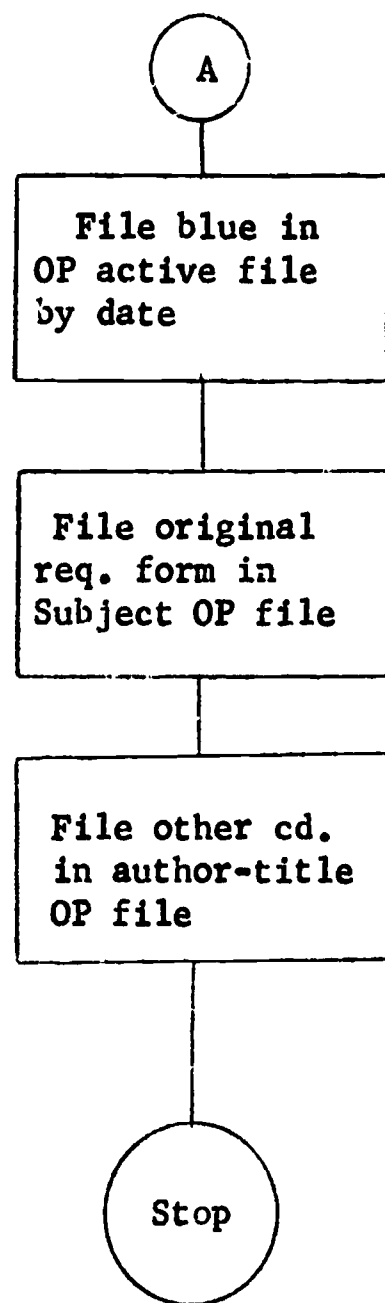
Library D--Acquisitions Department--Accounting



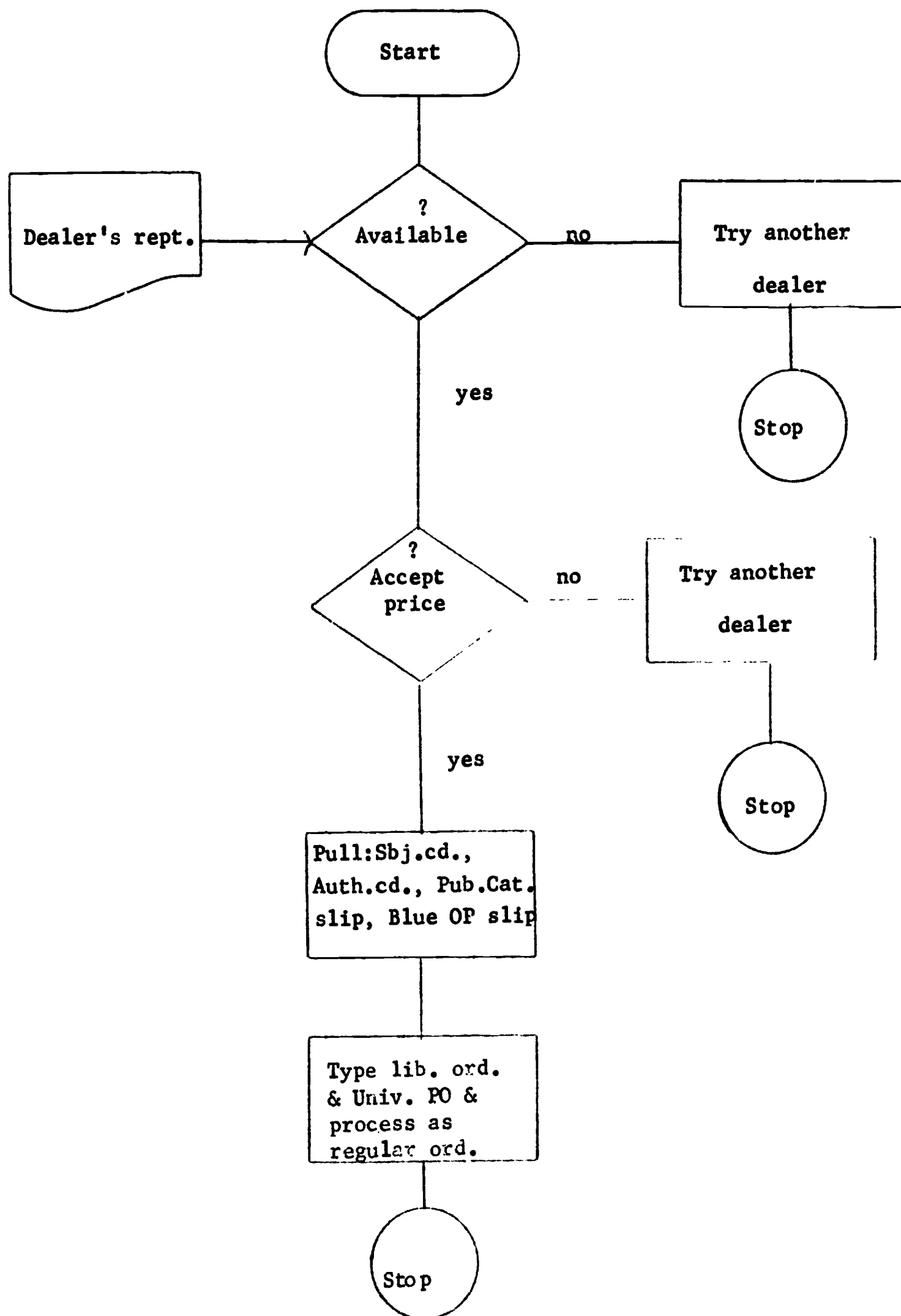
Library D--Acquisitions Department--OP

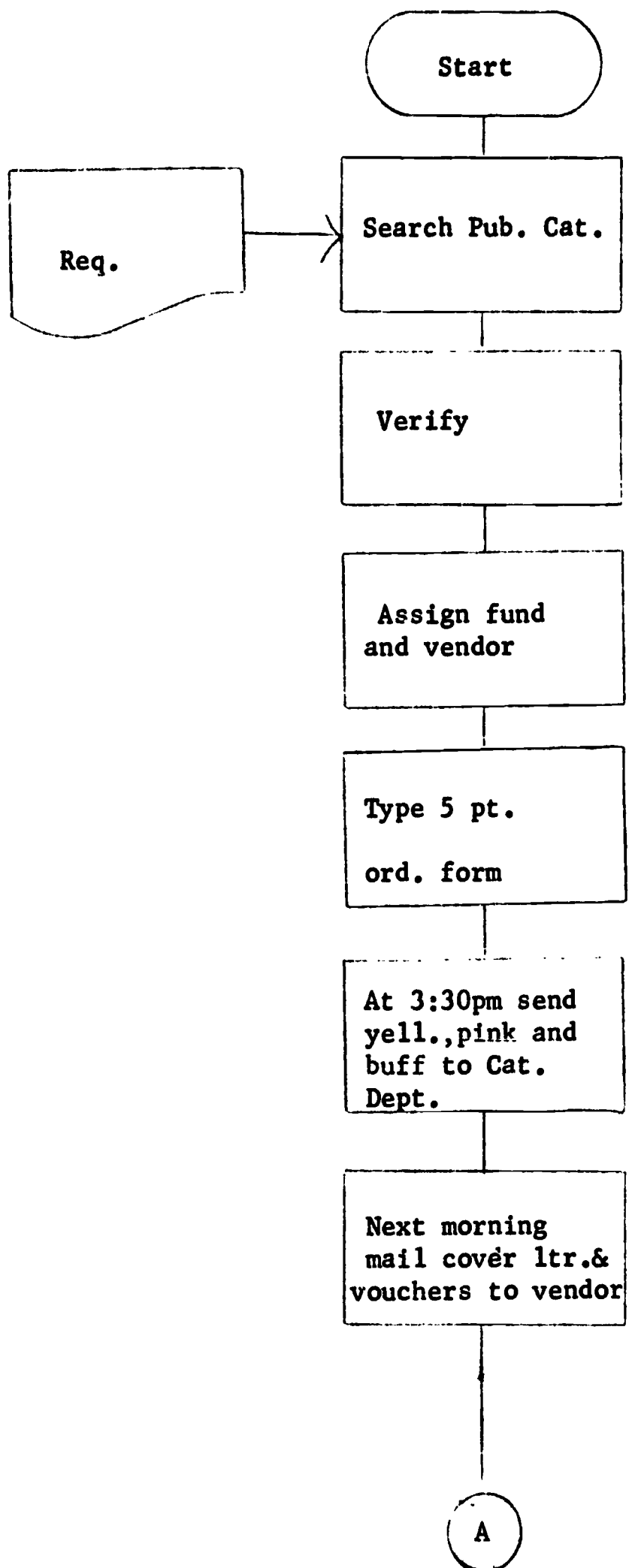


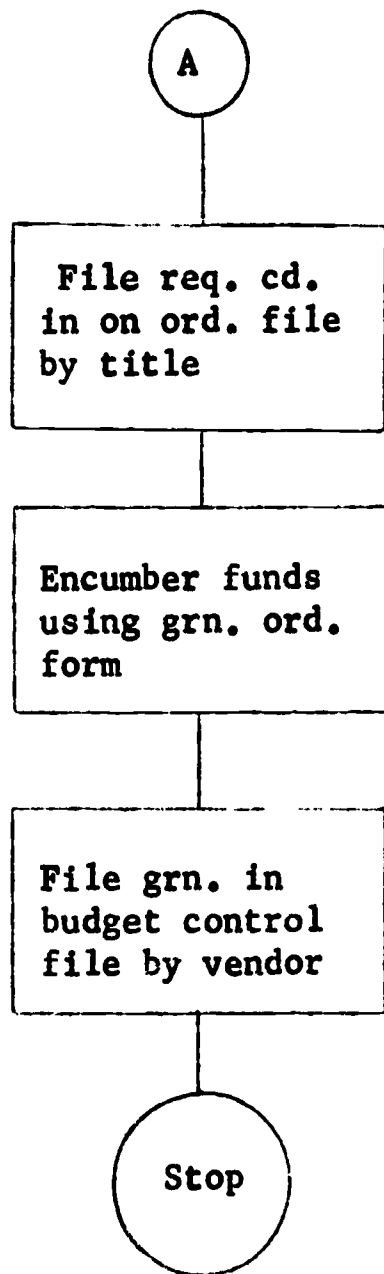
Library D--Acquisitions Department--OP



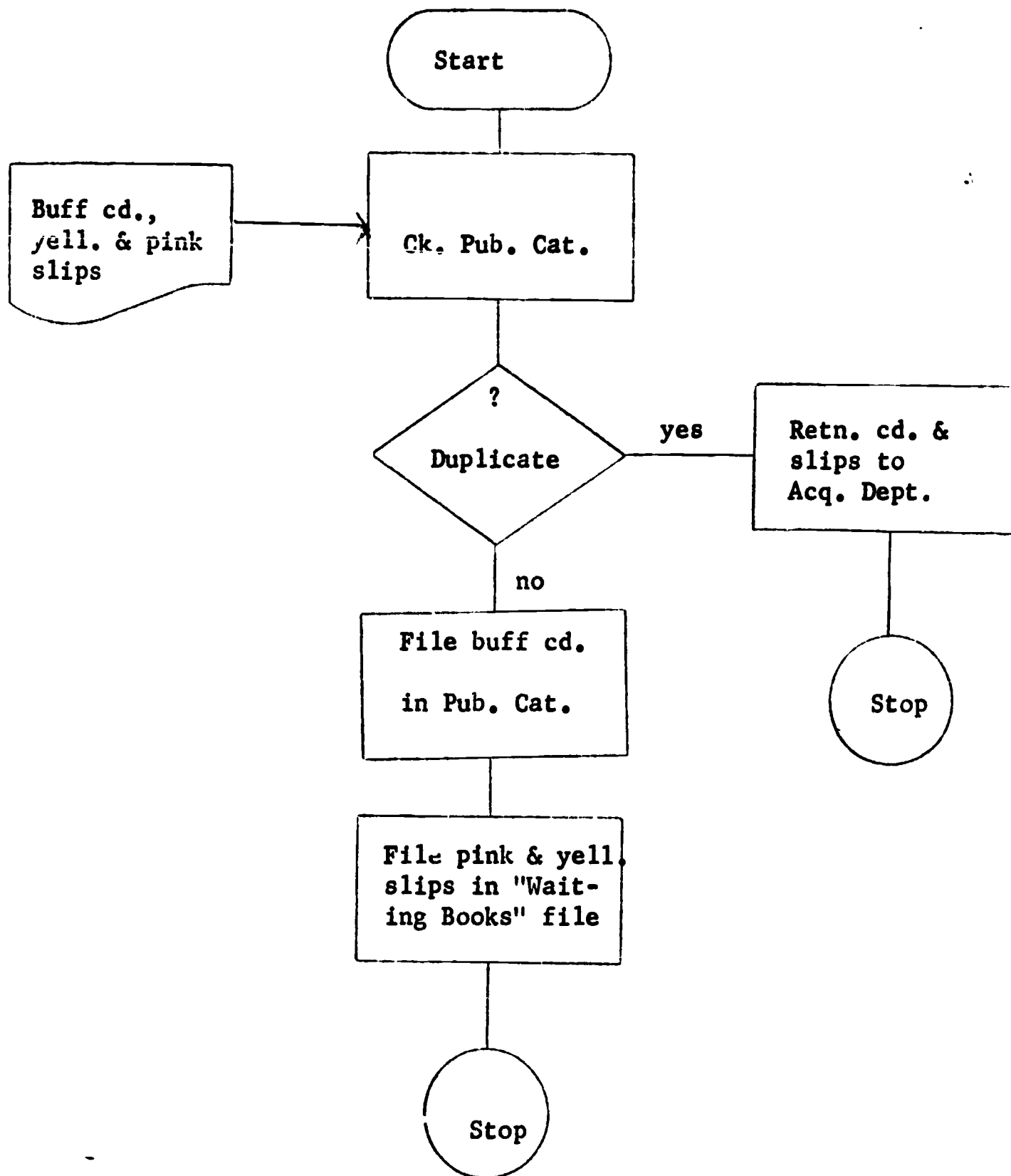
A - 50
Library D--Acquisitions Department--OP Report Processing



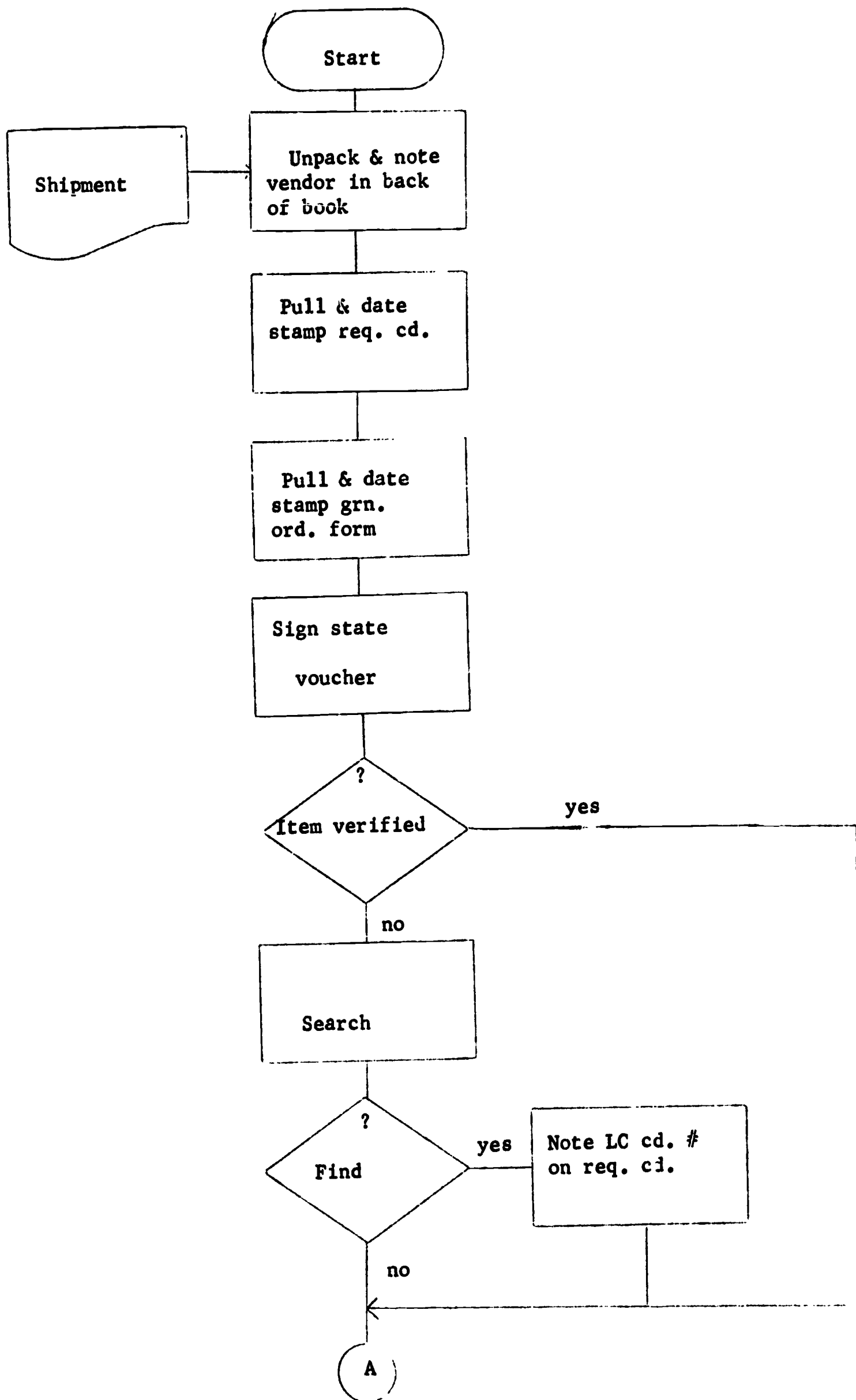


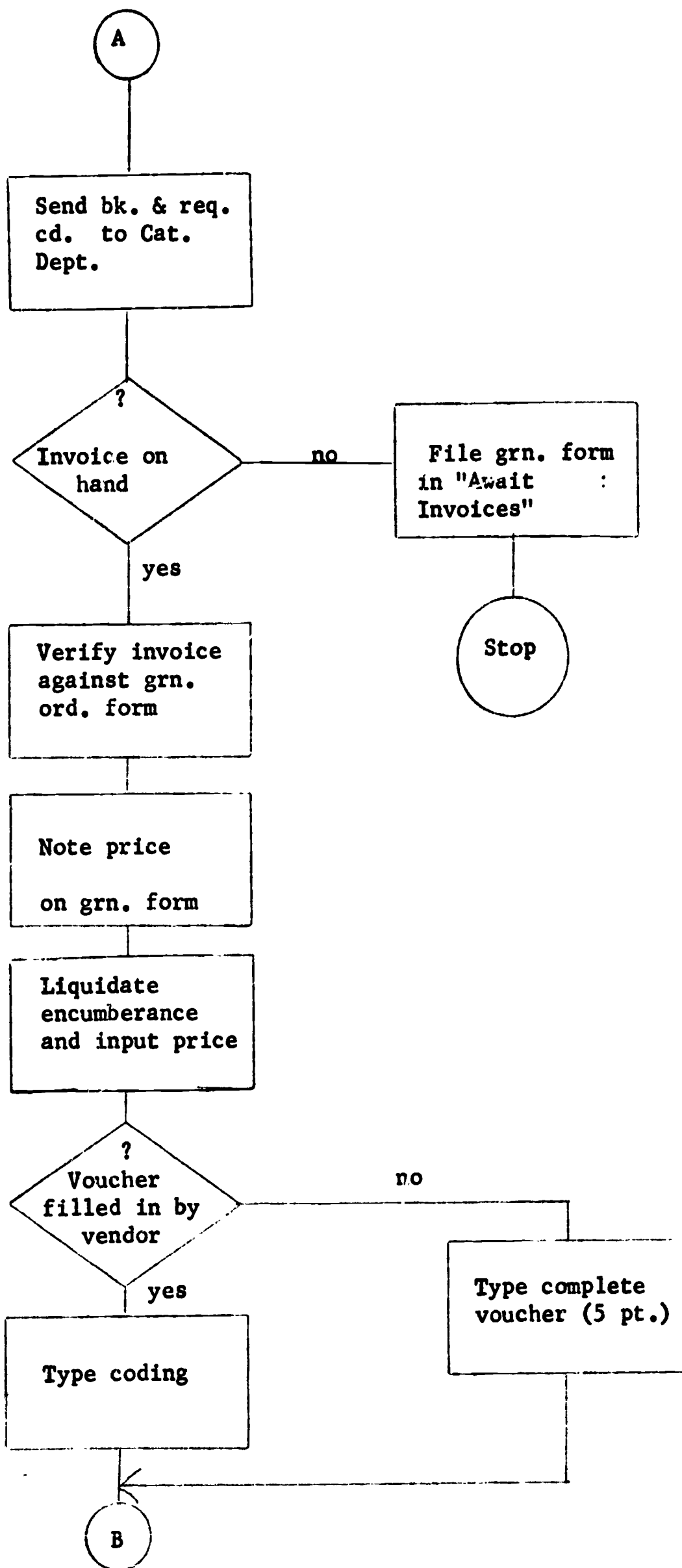


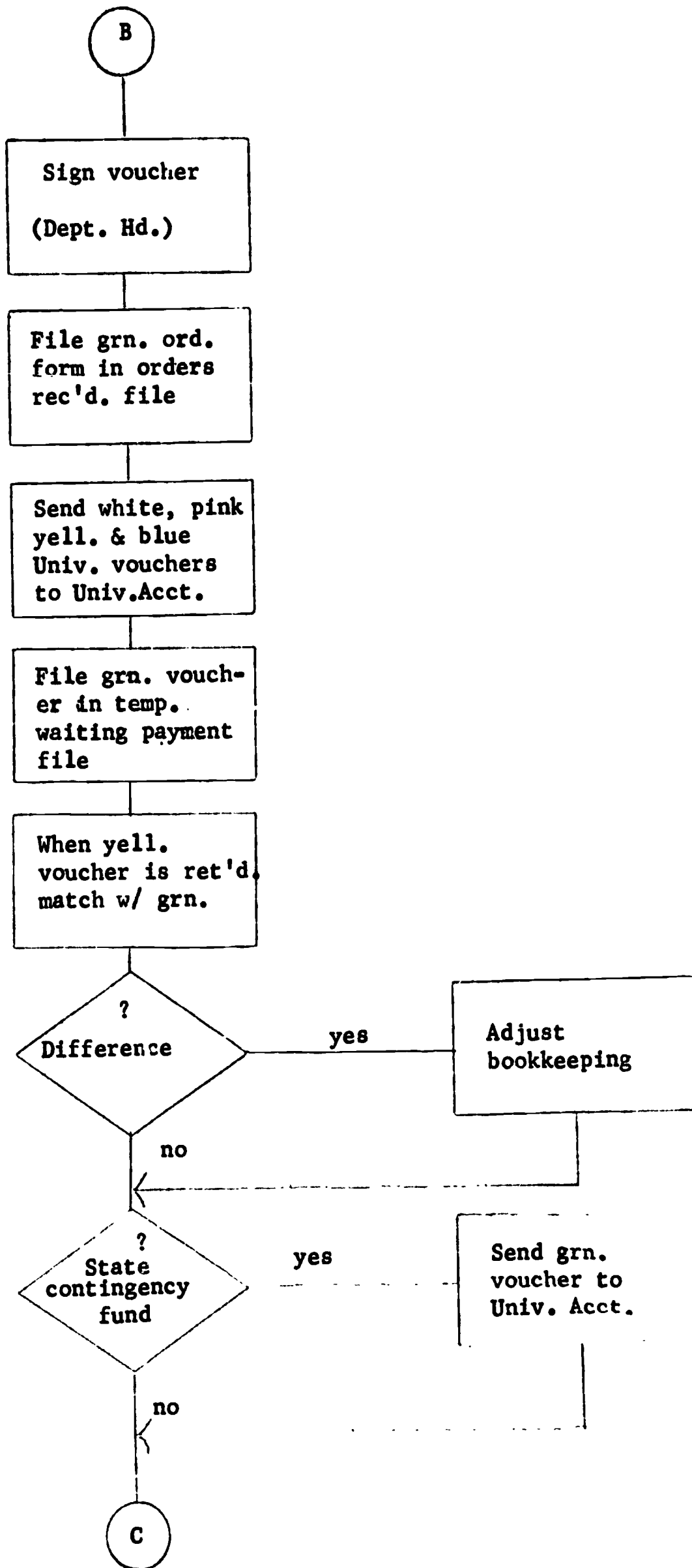
Library E--Catalog Department-- Ordering

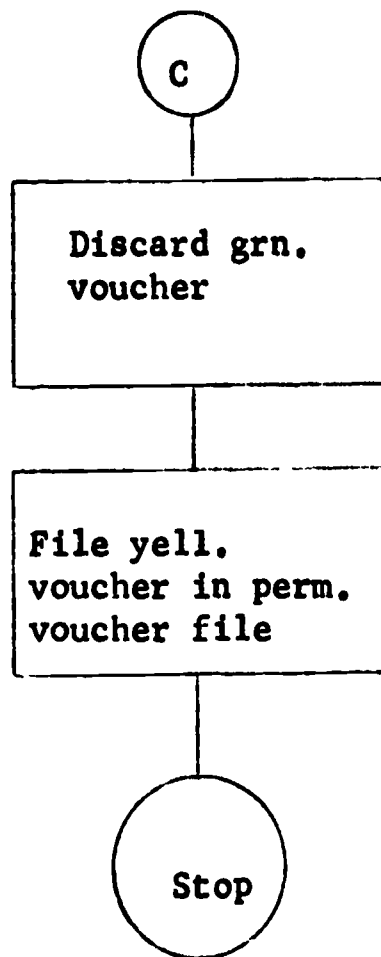


A - 50
Library E--Acquisitions Department--Receiving

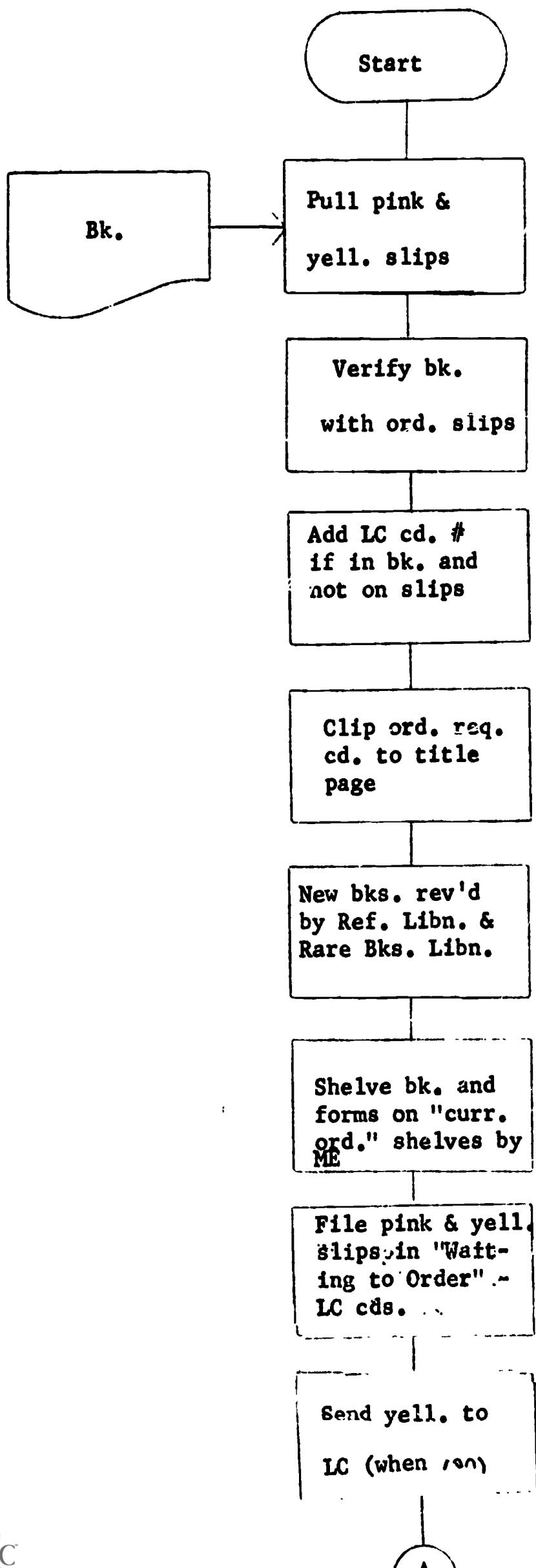




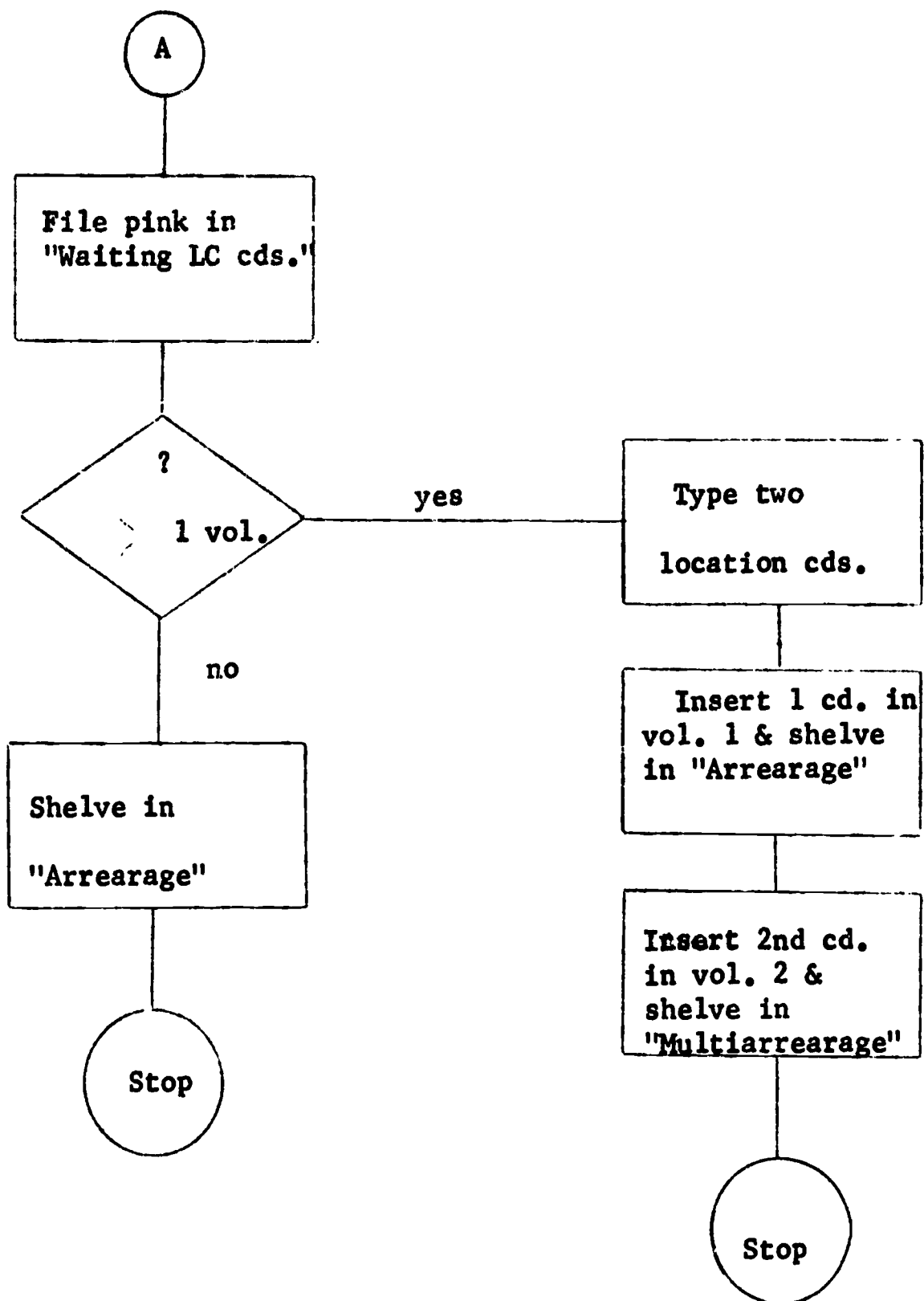




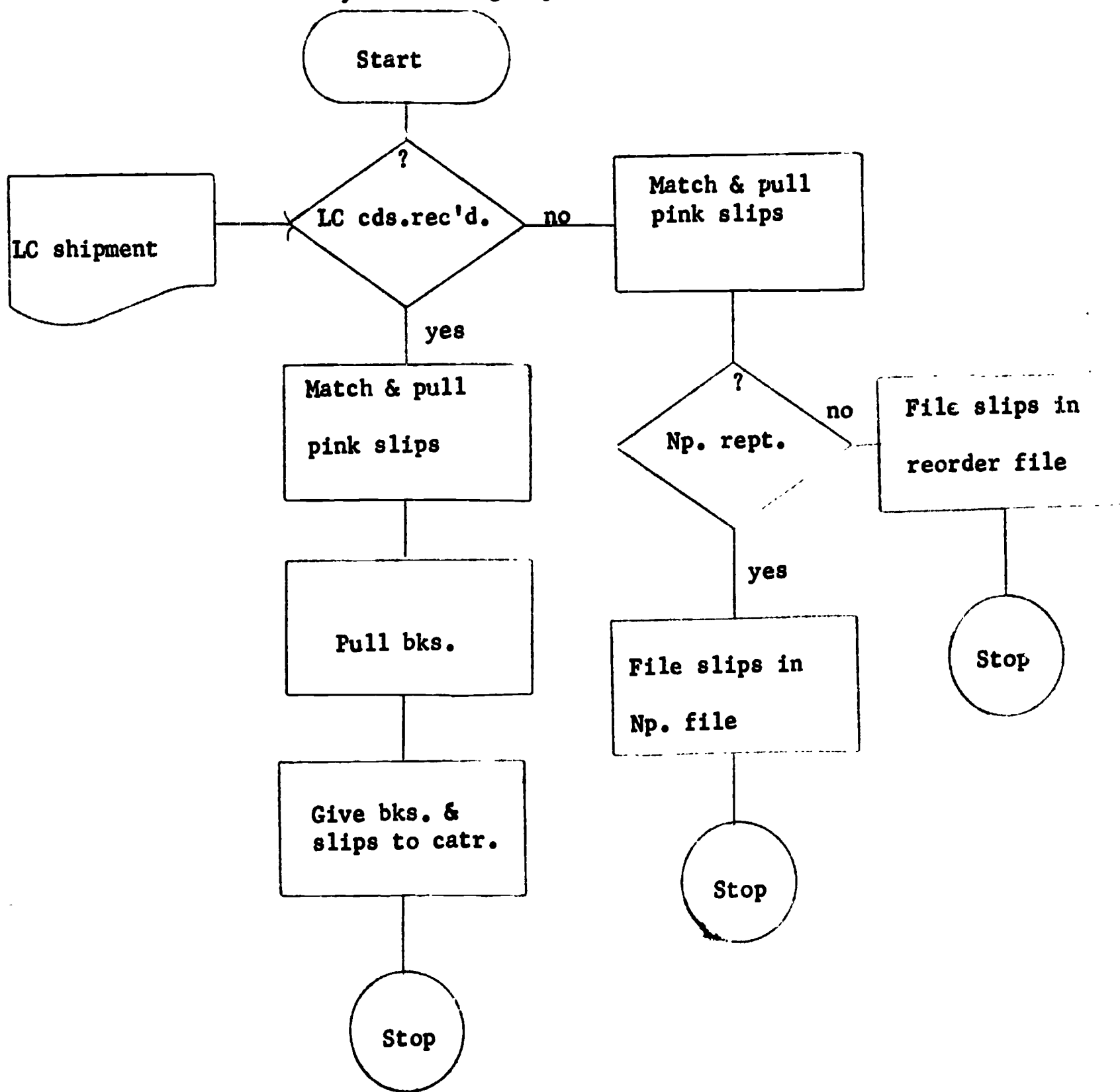
Library E--Catalog Department--LC Card Ordering



A - 65
Library E--Catalog Department--LC Card Ordering

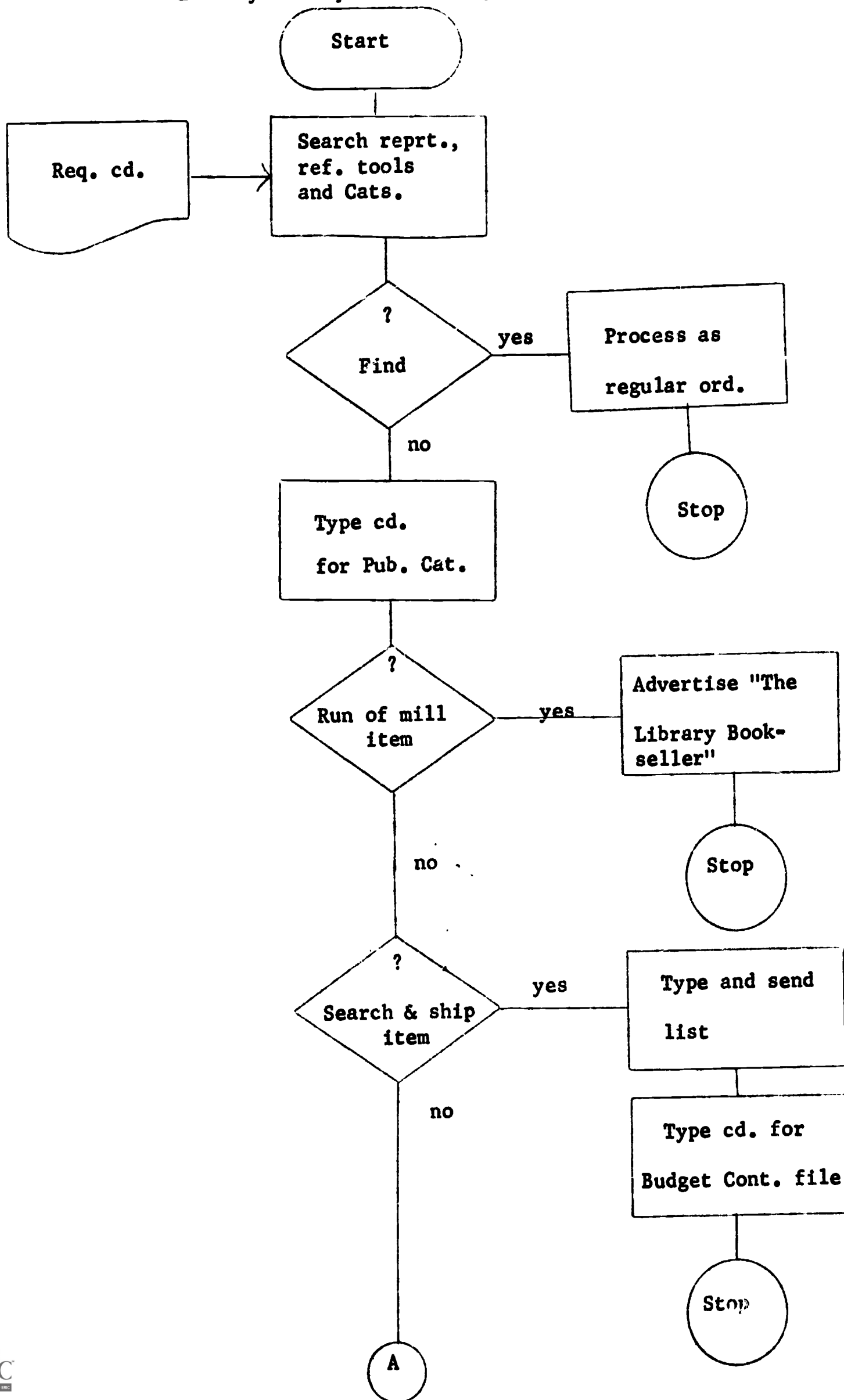


Library E--Catalog Department--LC Card Receiving

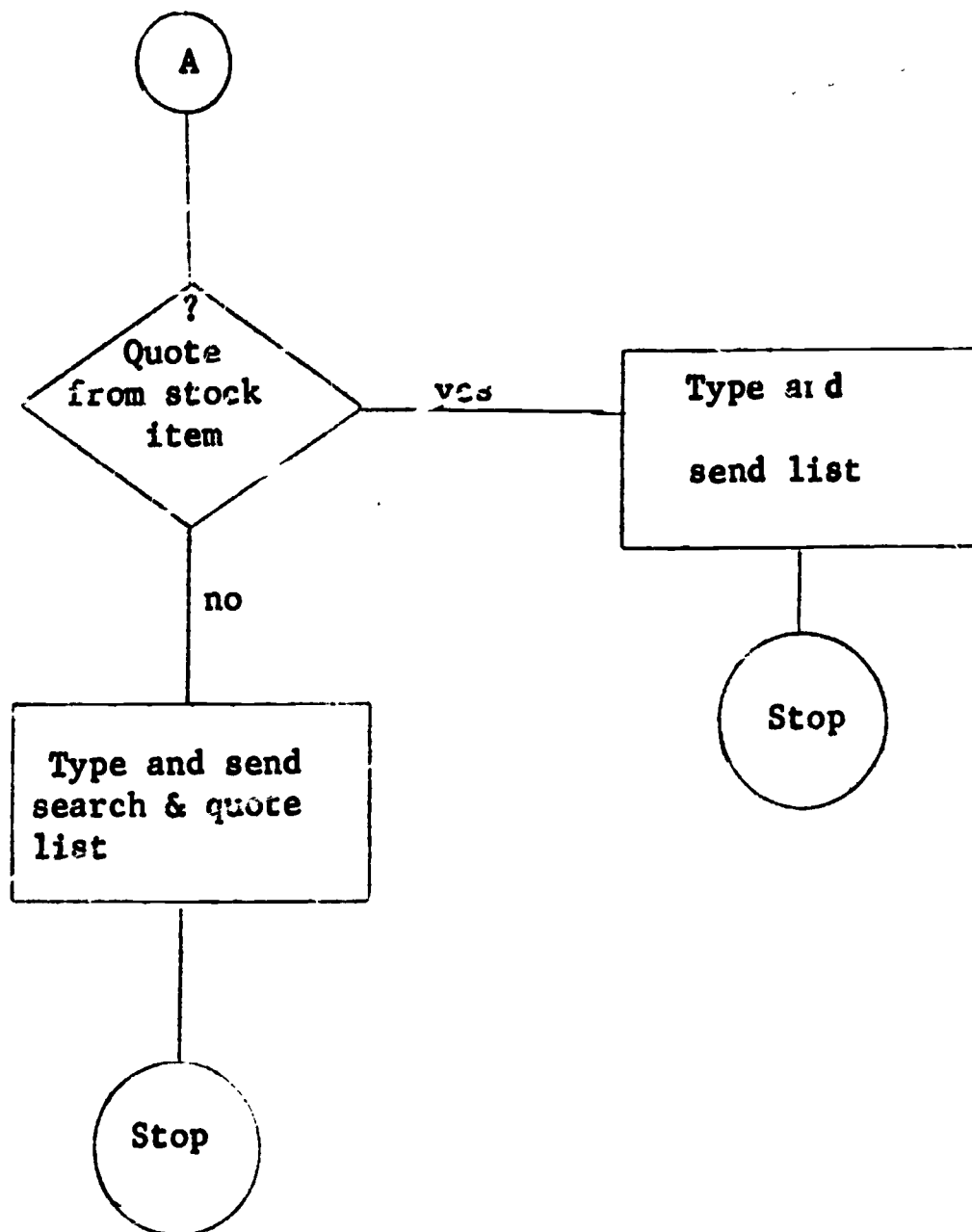


Library E--Acquisitions Department--Out of Print

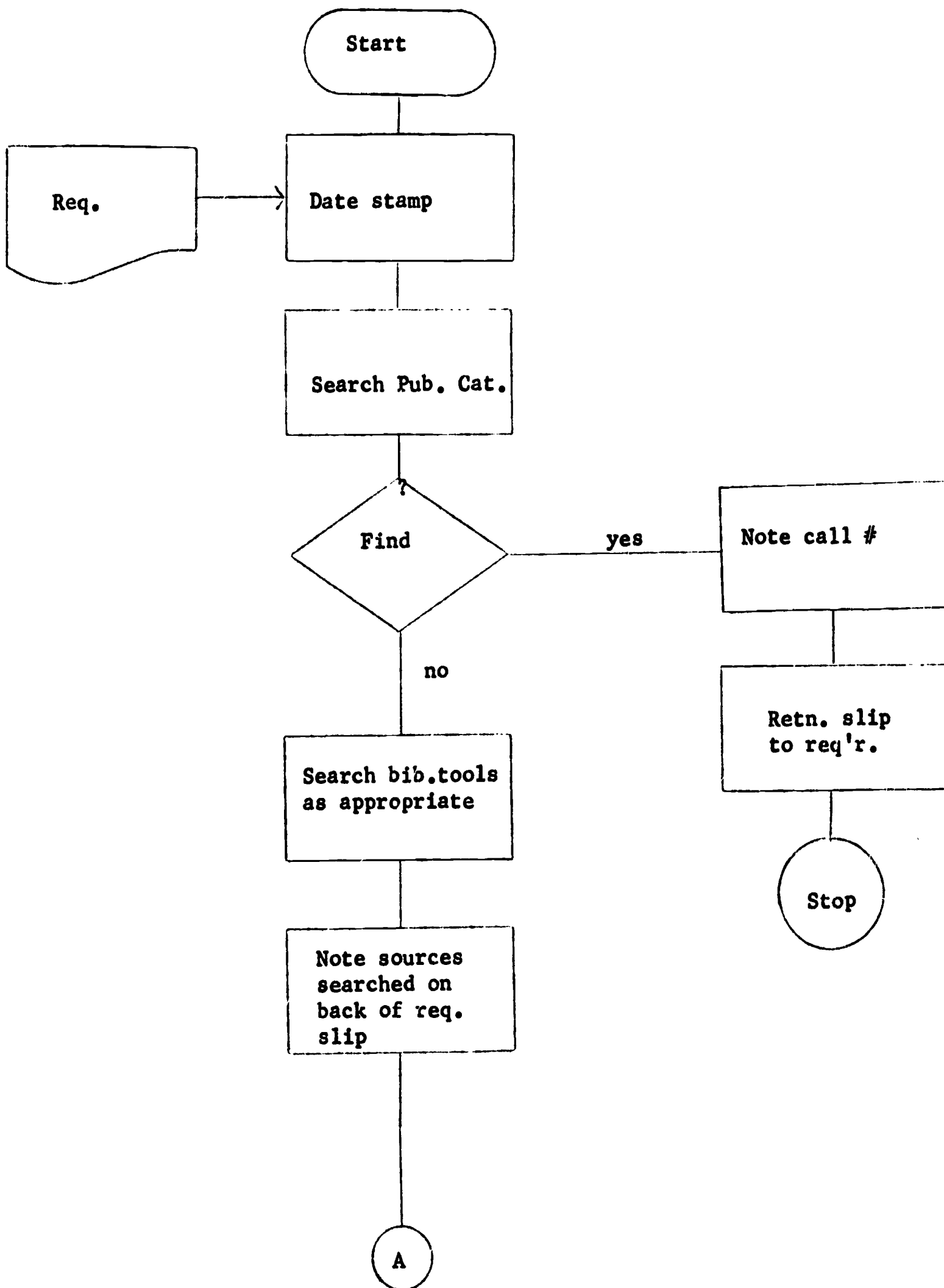
A - 67



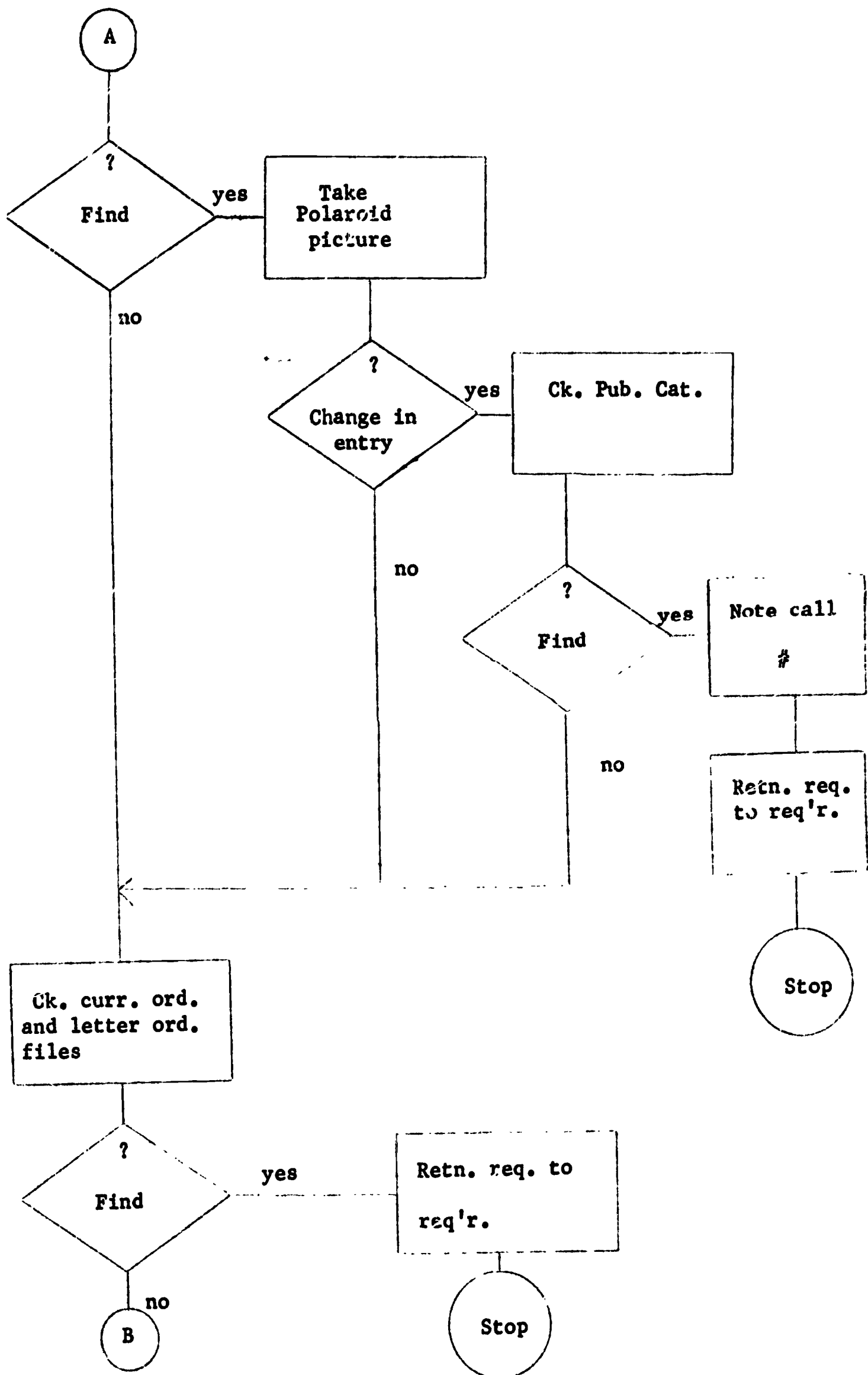
A - 63
Library E--Acquisitions Department--Out of Print



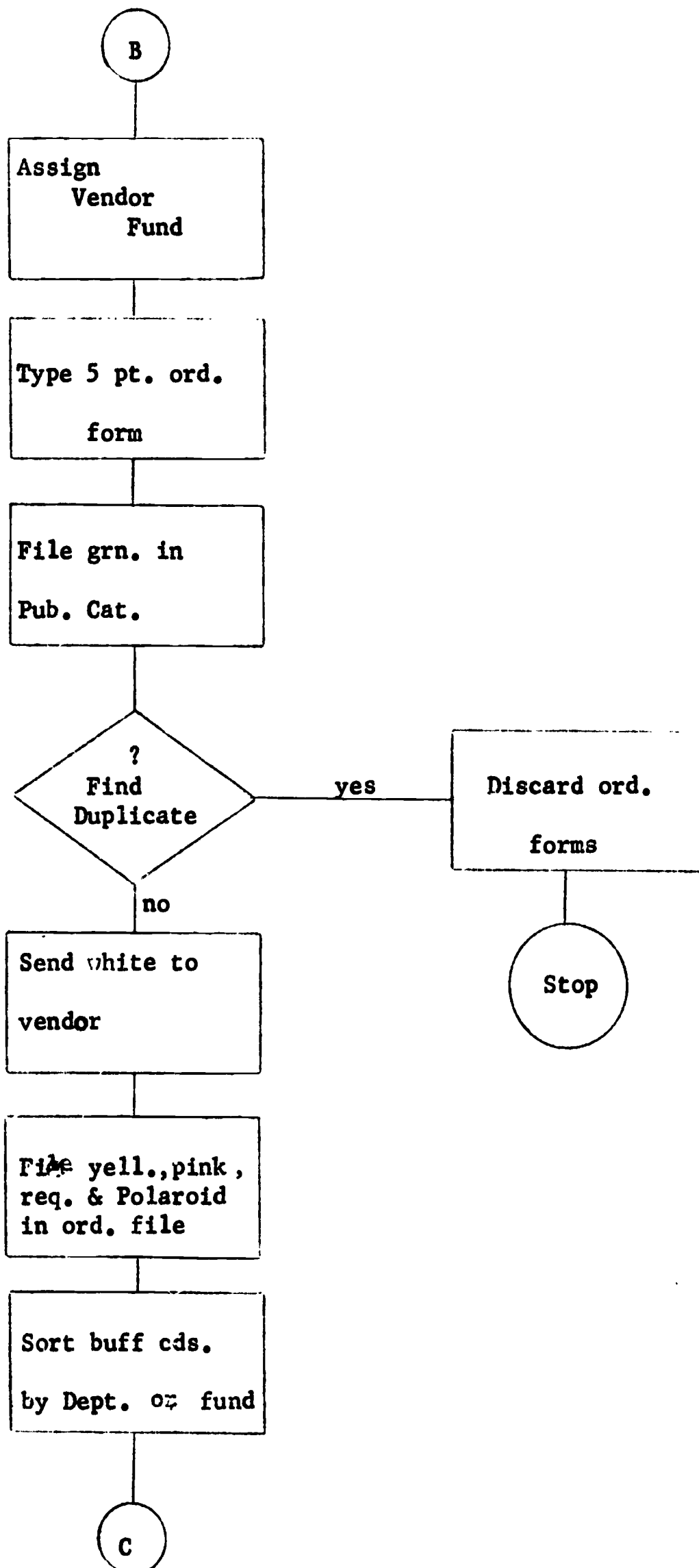
A - 69
Library F--Acquisitions Department--Ordering



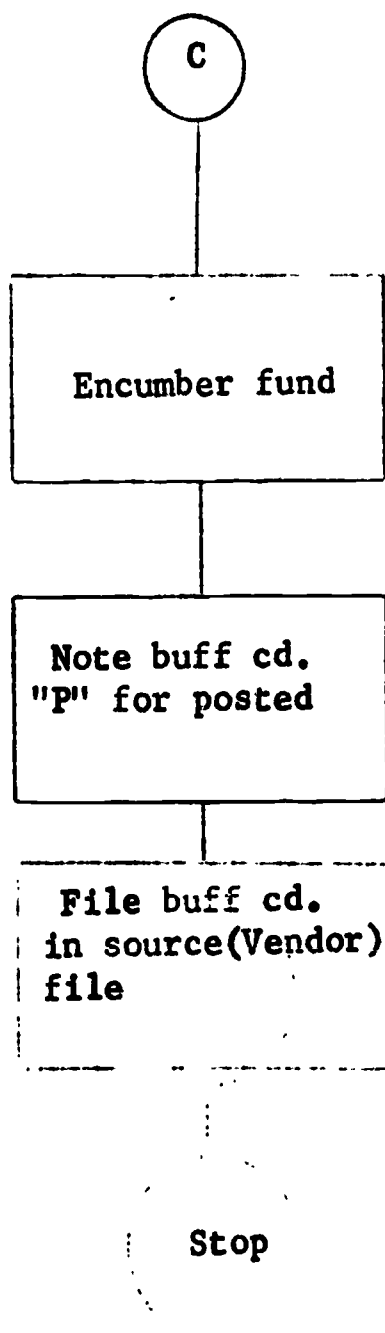
Library F--Acquisitions Department--Ordering



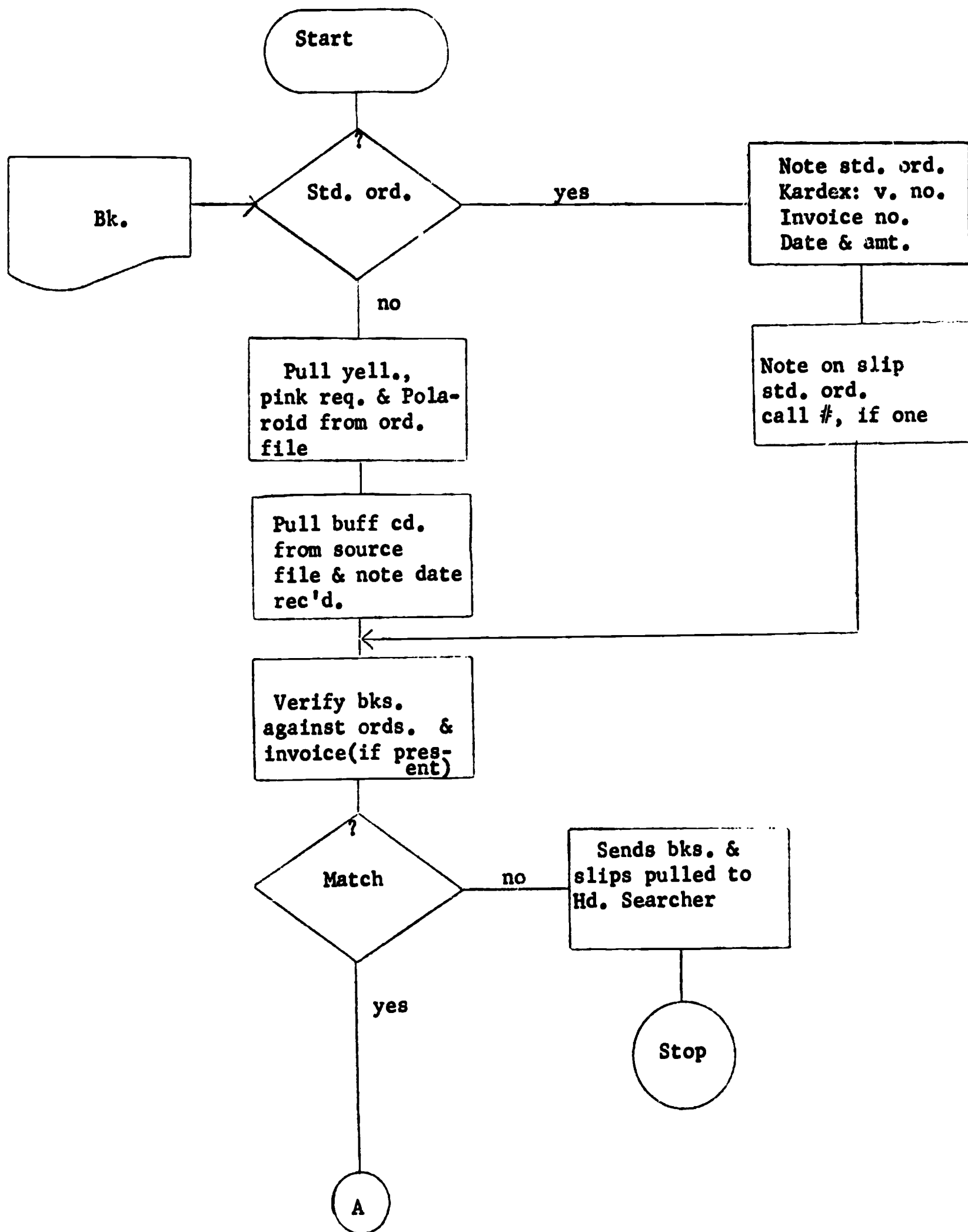
Library F--Acquisitions Department--Ordering



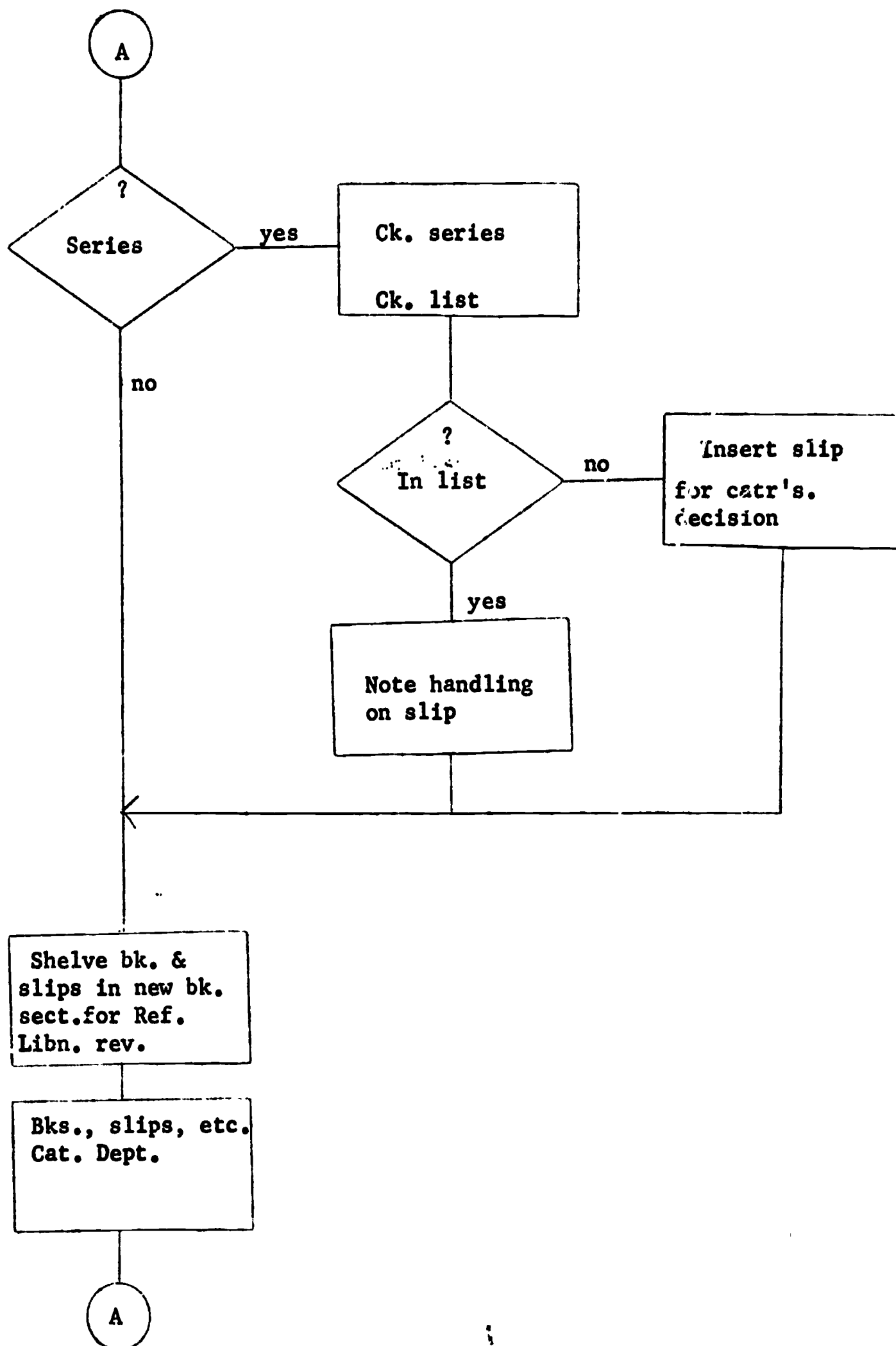
Library F--Acquisitions Department--Ordering



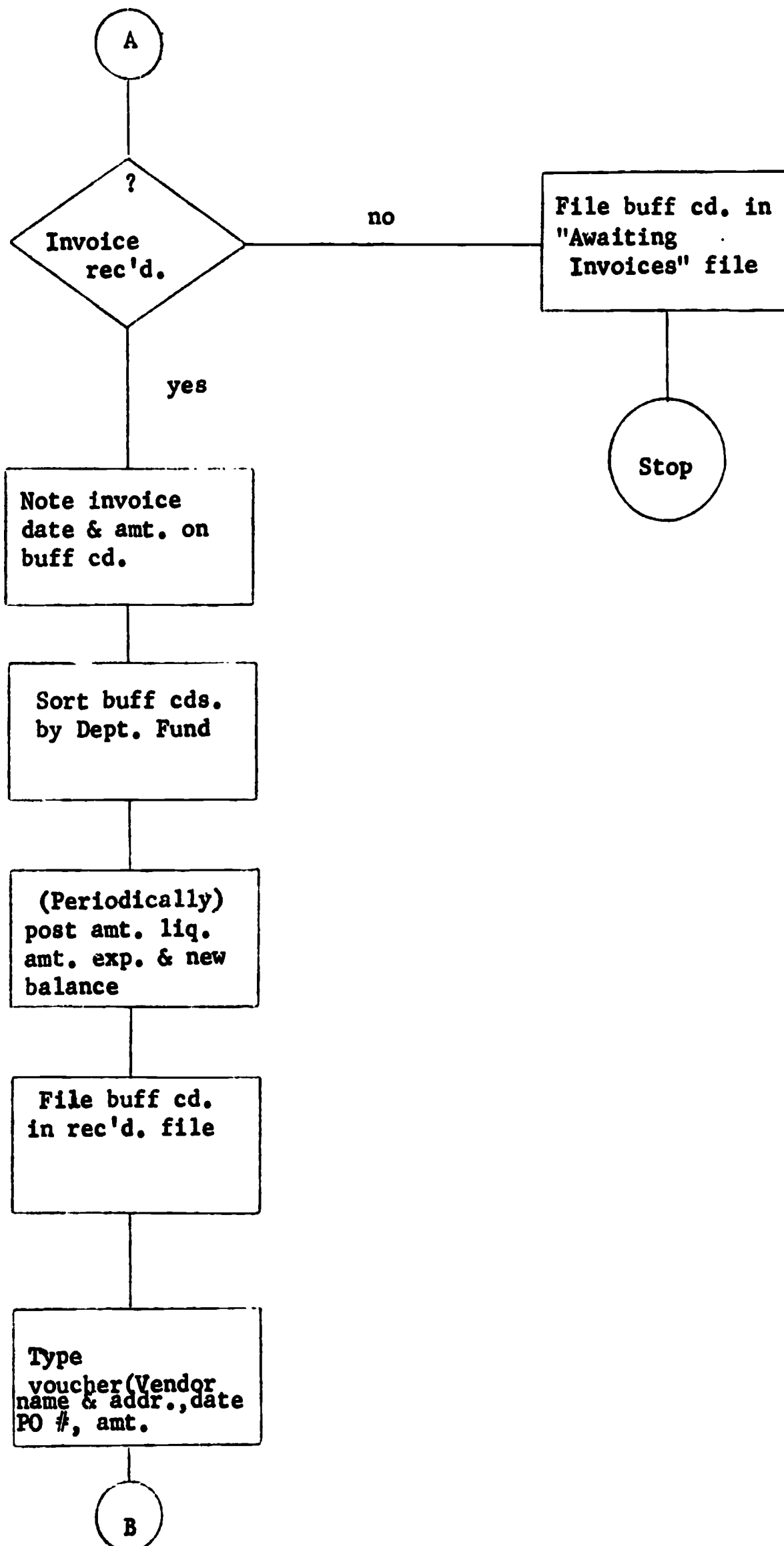
Library F--Acquisitions Department--Receiving



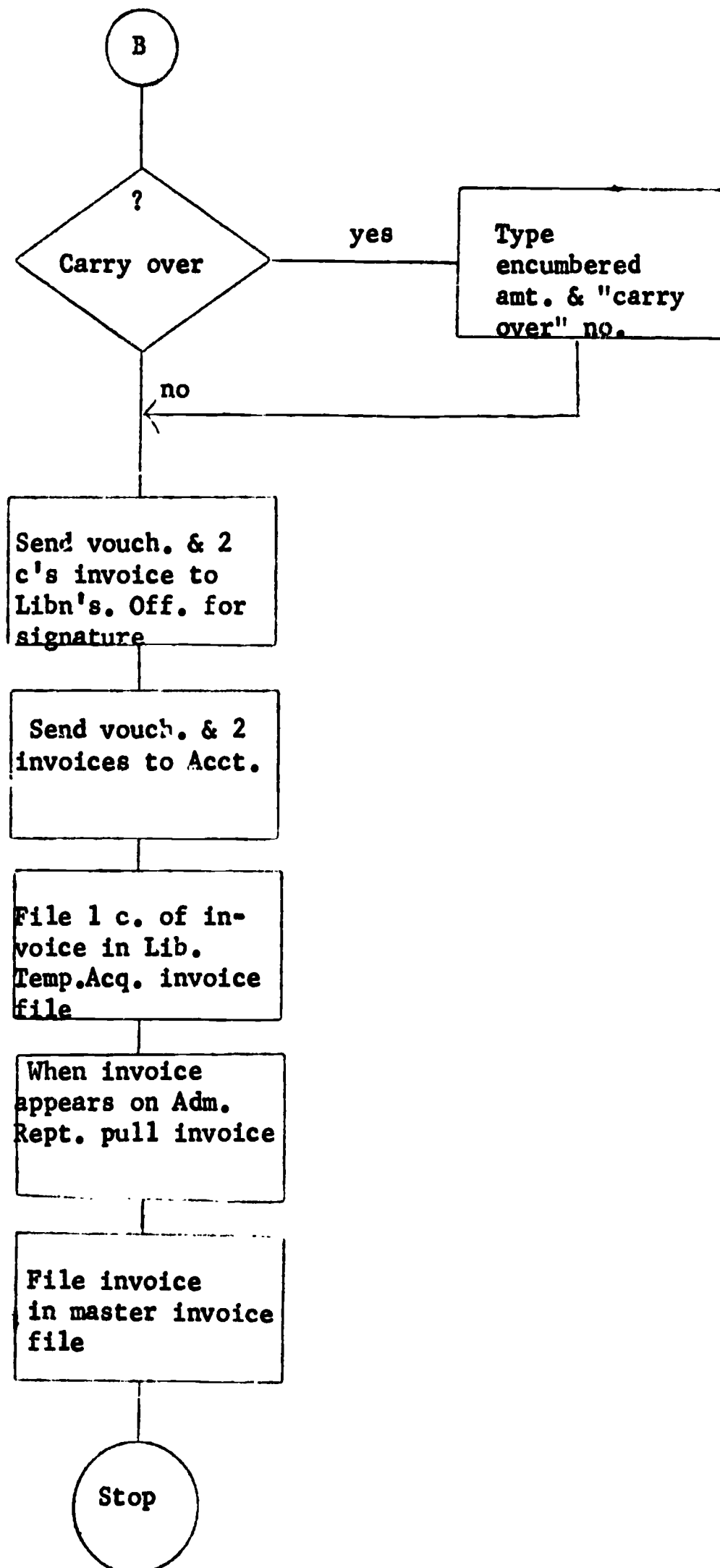
Library F--Acquisitions Department--Receiving



A - 75
Library F--Acquisitions Department--Receiving



A - 76
Library F--Acquisitions Department--Receiving



APPENDIX B

Current Operations Questionnaire

This questionnaire was compiled to learn the cost and promptness of present technical services operations preparatory to comparing them with those of the NELINET center. When all completed questionnaires have been returned, the data will be tabulated.

Present plans call for preparation of a journal article based on the data.

Dear

The enclosed questionnaire is designed to help the New England State University librarians assemble data on the efficiency and costs of current processing operations. In particular, data are being sought about those tasks for which the NELINET operation will soon offer alternatives, catalog card and book label production. The use of uniform criteria for these data is important not for the questionable comparisons of the six libraries that become possible, but because it will allow the evaluations of NELINET services to be compared.

The following are kinds of figures these data will make possible:

1. For every dollar spent for library materials,
x cents are spent for technical services salaries.
2. The average time from request of a recent American imprint until it is ready for shelving is:
ordering _____ days;
cataloging _____ days;
dealer _____ days;
total _____ days.
3. Cost per title for card production.
4. Cost per volume for label production.

Thank you for your cooperation.

Sincerely yours,

James E. Agenbroad
Library Systems Analyst

INSTRUCTIONS

If you do not wish your library to be identified in any subsequent publication of these data you may so specify.

Where information is unavailable or would require excessive labor for compilation, please give an estimate and indicate it by a "*".

Please return one completed copy of the questionnaire with the used work sheets and list of totals obtained in section IX to Inforonics by February 15, 1968.

The second copy of the questionnaire is for your records.

Sections I-V

It is extremely desirable that salary, book fund and book collection data have the same basis. Thus, if technical services are provided for state colleges whose book funds and collection size statistics are not included, salaries for these people should be excluded; if a branch library does its own ordering and/or cataloging, exclude their book collection and book fund data unless salaries of the branch processing staff are included with those of the main library; and if reclassification constitutes a significant proportion of the catalog dept. operation, the salaries spent for it should be excluded since reclassification is not part of the book fund or collection growth data. Please use data for the year ending June 30, 1967.

Sections VIII-IX

In these sections data on card and label production for reclassification and for state college processing may be included or not as is more convenient as long as the same procedure is followed for all three kinds of data: supplies, labor times, and production quantity.

Extra work sheets are available.

Library

I. EXPENSES

I.1. Total library expenses (staff, books, equipment and supplies) for 7/66-6/67: _____

I.2.1. Total funds for library materials: _____

I.2.2. If this included binding, how much for binding? _____

I.2.3. If this included subscription renewals, how much for subscription renewals? _____

I.2.4. If this included a fund for new subscriptions, how much for new subscriptions? _____

I.3.1. How much for new furniture and equipment? Total library _____, catalog dept. _____, order dept. _____.

I.3.2. How much for maintenance and repair of furniture and equipment? Total library _____, catalog dept. _____, order dept. _____.

I.3.3. How much for leased equipment? Total library _____, catalog dept. _____, order dept. _____.

I.4. How much for supplies? Total library _____, catalog dept. _____, order dept. _____.

II. PHYSICAL PLANT

Estimated replacement cost of:

- a. Library with entire contents _____
- b. All equipment and furniture _____
- c. Catalog dept. equipment and furniture _____
- d. Order dept. equipment and furniture _____
- e. Card and label production equipment and furniture _____

III. PERSONNEL AND SALARIES

III.1. Total Library staff:

| | | |
|------------|-----------------|------------|
| Librarians | _____ man years | _____ cost |
| Clerical | _____ man years | _____ cost |
| Student | _____ hours | _____ cost |

III.2. Catalog Dept. Staff

Librarians _____ man years _____ cost
 Clerical _____ man years _____ cost
 Student _____ hours _____ cost

III.3. Order Dept. Staff (Please include bookkeeping even when done elsewhere in the library)

Librarians _____ man years _____ cost
 Clerical _____ man years _____ cost
 Student _____ hours _____ cost

III.4. Head of technical processing and immediate staff:

Librarians _____ man years _____ cost
 Clerical _____ man years _____ cost
 Student _____ hours _____ cost

III.5. Other technical processing staff (binding, serials, reclassification, archives, maps, govt. docs., processing for state colleges, etc.)

Librarians _____ man years _____ cost
 Clerical _____ man years _____ cost
 Student _____ hours _____ cost

What areas does this include?

III.6. Administrative staff of library:

Librarians _____ man years _____ cost
 Clerical _____ man years _____ cost
 Student _____ hours _____ cost

III.7. Hours per week: librarians _____ clerks _____
 Weeks per year: librarians _____ clerks _____
 Paid holidays per year _____.

IV. COLLECTION SIZE

1. Total in library on June 30, 1967 _____ vols; _____ titles
2. Does this include govt. docs? _____
3. Does this include microforms? _____
4. Serials _____ vols. _____ titles

V. COLLECTION GROWTH 7/66-6/67

1. Added _____ vols; _____ titles;
2. Cataloged by staff in III.2. _____ vols; _____ titles;

If V.1 and 2 differ, why? (Uncataloged documents, gifts, etc.)

3. Cataloged for branches _____ vols. _____ titles.
4. Serials added _____ vols. _____ titles.
5. Gifts added _____ vols. _____ titles.
6. Reclassified _____ vols. _____ titles.

VI. SPACE

VI. Square feet devoted to the following: Catalog
Dept. _____; Order Dept. _____; Administrative _____;
Stack space _____; All staff work space _____; Total
library _____.

VII. OPERATION TIMES

VII.1. Select randomly from books going from
processing to the stacks 30 recent American trade titles and
supply as much of the following data for each as possible.

| | A. Date requested by faculty | B. Date order sent out | C. Date book rec'd or paid for | D. Date sent to cat. dept. | E. Date sent to stacks |
|-----|------------------------------------|------------------------------|--------------------------------------|----------------------------------|------------------------------|
| 1. | _____ | _____ | _____ | _____ | _____ |
| 2. | _____ | _____ | _____ | _____ | _____ |
| 3. | _____ | _____ | _____ | _____ | _____ |
| 4. | _____ | _____ | _____ | _____ | _____ |
| 5. | _____ | _____ | _____ | _____ | _____ |
| 6. | _____ | _____ | _____ | _____ | _____ |
| 7. | _____ | _____ | _____ | _____ | _____ |
| 8. | _____ | _____ | _____ | _____ | _____ |
| 9. | _____ | _____ | _____ | _____ | _____ |
| 10. | _____ | _____ | _____ | _____ | _____ |
| 11. | _____ | _____ | _____ | _____ | _____ |
| 12. | _____ | _____ | _____ | _____ | _____ |
| 13. | _____ | _____ | _____ | _____ | _____ |
| 14. | _____ | _____ | _____ | _____ | _____ |
| 15. | _____ | _____ | _____ | _____ | _____ |
| 16. | _____ | _____ | _____ | _____ | _____ |

| | A | B | C | D | E |
|-----|---|---|---|---|---|
| 17. | | | | | |
| 18. | | | | | |
| 19. | | | | | |
| 20. | | | | | |
| 21. | | | | | |
| 22. | | | | | |
| 23. | | | | | |
| 24. | | | | | |
| 25. | | | | | |
| 26. | | | | | |
| 27. | | | | | |
| 28. | | | | | |
| 29. | | | | | |
| 30. | | | | | |

VII. 2. Select randomly from cards just typed and about to be arranged for filing into the catalog 30 main entries for recent American trade titles and supply as much of the following data as possible:

| | A. Date requested by faculty | B. Date order sent out | C. Date book rec'd or paid for | D. Date sent to cat. dept. | E. Date of this test |
|-----|------------------------------|------------------------|--------------------------------|----------------------------|----------------------|
| 1. | | | | | |
| 2. | | | | | |
| 3. | | | | | |
| 4. | | | | | |
| 5. | | | | | |
| 6. | | | | | |
| 7. | | | | | |
| 8. | | | | | |
| 9. | | | | | |
| 10. | | | | | |
| 11. | | | | | |
| 12. | | | | | |
| 13. | | | | | |
| 14. | | | | | |

B. - 8

| | A | B | C | D | E |
|-----|-------|-------|-------|-------|-------|
| 14. | _____ | _____ | _____ | _____ | _____ |
| 15. | _____ | _____ | _____ | _____ | _____ |
| 16. | _____ | _____ | _____ | _____ | _____ |
| 17. | _____ | _____ | _____ | _____ | _____ |
| 18. | _____ | _____ | _____ | _____ | _____ |
| 19. | _____ | _____ | _____ | _____ | _____ |
| 20. | _____ | _____ | _____ | _____ | _____ |
| 21. | _____ | _____ | _____ | _____ | _____ |
| 22. | _____ | _____ | _____ | _____ | _____ |
| 23. | _____ | _____ | _____ | _____ | _____ |
| 24. | _____ | _____ | _____ | _____ | _____ |
| 25. | _____ | _____ | _____ | _____ | _____ |
| 26. | _____ | _____ | _____ | _____ | _____ |
| 27. | _____ | _____ | _____ | _____ | _____ |
| 28. | _____ | _____ | _____ | _____ | _____ |
| 29. | _____ | _____ | _____ | _____ | _____ |
| 30. | _____ | _____ | _____ | _____ | _____ |

VIII. SUPPLY COSTS

VIII. List actual or estimated supply costs for card and label production during 7/66-6/67. Please add items omitted from the list. Where supplies are also used for other purposes (typewriter maintenance for instance) estimate the proportion used for card and label production.

1. Card stock _____
2. L.C. cards _____
3. L.C. proof slips _____
4. Pressure sensitive labels _____
5. Polaroid film _____
6. Xerox supplies _____
7. Xerox maintenance _____
8. Flexowriter maintenance _____
9. Flexowriter supplies _____
10. Typewriter maintenance _____
11. Selin supplies _____

- 12. Xerox rental _____
- 13. Snopake _____
- 14. Other, specify _____
- 15. _____
- 16. _____

IX. LABOR TIME, COSTS, AND PRODUCTION

IX.1.1. Procedure: 1. For one "normal" week, have all staff involved in production of cards and labels fill out the daily work sheets indicating the minutes they spend at each task. It should be stressed that all the time devoted to producing cards and labels is to be included but that the time spent using the final product (such as filing catalog cards and pasting labels) should be excluded.

IX.1.2. At the end of the week arrange the sheets according to the salary of the workers.

IX.1.3. Add together the minutes spent at each task by all individuals earning the same salary and working the same hours/week/year.

IX.1.4. List these totals on a separate sheet in the format:

| TASK | MINUTES | SALARY | PER |
|------|---------|--------|-----|
| A1 | 360 | 1.75 | hr. |
| A1 | 2400 | 3,000 | yr. |

IX.2. In a locally convenient manner, obtain data on card sets produced and volumes labelled during the above week: _____ sets; _____ vols.

X. L.C. COPY

Proportion of acquisitions for which L.C. cataloging (L.C. cards, proof slips, title II cards, PW, or LC entries in NUC) is not used (not available or too slow).

_____ If L.C. cards are ordered, how long until receipt of the bulk of an order? _____

XI. OVERHEAD

XI.1. If any of the following are excluded from the library expenses in I. above but are incurred by operation of the library, please list or estimate below the amount spent during 7/66-6/67. (For instance, if the president, vice president and treasurer devote one half their time to fund raising and four per cent of the funds raised are applied to library operation, then two per cent of their salaries are spent to acquire operating funds for the library.)

XI.2. If a figure used by the university to compute the overhead costs on applications for federal research grants (at one university an additional 40% of the salaries of the researchers) is an accurate indicator of the library overhead (covers all the items listed below) write it here and explain its use: _____

If such a figure exists, but does not reflect library overhead, write it here; explain its use and then give data for items not covered below. _____

XI.3. Fringe benefits to library staff (if overall fringe benefits per person or per faculty and per clerical figures are available just supply them.)

- a. University retirement contribution _____
- b. University health insurance contribution _____
- c. University life insurance contribution _____
- d. Travel expenses paid to staff _____
- e. Relocation grants to new staff _____
- f. Tuition grants to staff and their families _____
- g. Staff use of University subsidized clinic _____
- h. Staff use of University subsidized housing _____
- i. Other, please specify _____

XI.4. Utilities:

- a. Heat
- b. Air conditioning
- c. Electricity
- d. Water
- e. Telephone
- f. Postage
- g. Other, specify _____

XI 5. Services:

- a. Messenger
- b. Payroll and Billing
- c. Accounting and Auditing
- d. Janitorial
- e. Architectural and Interior Decorating
- f. Equipment purchase handling
- g. Printing
- h. Grounds
- i. Repairs to building and equipment
- j. Data processing, programming
and computer time
- k. Personnel services such as typing tests
for clerks
- l. Fund raising by alumni office and
university administration
- m. Proportionate share of general university
overhead such as news bureau, fire
insurance, etc.
- n. Other, specify _____

XII. QUESTIONNAIRE

XII.1. Estimated time and cost of completing this
questionnaire. _____

XII.2. Comments on this questionnaire.

Circle One: librarian clerk student Salary: \$ _____ per
 Date _____ Hours per week worked _____; weeks per year worked _____

INSTRUCTIONS:

Write the minutes you work at any task during each hour in the box for it.

ACQUIRE COPY

A1

A2

A3

A4

A5

A6

PREPARE COPY

P1

P2

P3

P4

P5

P6

P7

REPRODUCTION

R1

R2

R3

R4

| | | | | | | | | | | | | | | | | | | | |
|-------------|-------------------------------------|---------------------------------------|-------------------|--------------------|-----------------------------------|---------------------------|---------------------------------------|--------------|-------------------------------------|----------------------------|--------------------------------|----------------------------------|---------------------------|-------------------------------------|-----------------|-------------------------------------|-----------------------|--------------------------|-----------|
| Hour | Filing proof slips or Title 2 cards | Searching NUC, proof slip or T2 files | Ordering LC cards | Receiving LC cards | Polaroid camera use & maintenance | Supervise & train A1 - A5 | Modify LC card, slip or Polaroid copy | Trim & paste | Type call number on LC card or slip | Type & edit card for Xerox | Arrange cards for reproduction | Type & edit copy for Flexowriter | Supervise & train P1 - P6 | Xerox maintenance, load clean, etc. | Xerox operation | Flexowriter maintenance, load, etc. | Flexowriter operation | Separate & arrange cards | Supervise |
| 8 - 9 | | | | | | | | | | | | | | | | | | | |
| 9 - 10 | | | | | | | | | | | | | | | | | | | |
| 10 - 11 | | | | | | | | | | | | | | | | | | | |
| 11 - 12 | | | | | | | | | | | | | | | | | | | |
| 1 - 2 | | | | | | | | | | | | | | | | | | | |
| 2 - 3 | | | | | | | | | | | | | | | | | | | |
| 3 - 4 | | | | | | | | | | | | | | | | | | | |
| 4 - 5 | | | | | | | | | | | | | | | | | | | |
| other hours | | | | | | | | | | | | | | | | | | | |
| Total | | | | | | | | | | | | | | | | | | | |

Salary: \$ _____ per
 ; weeks per year worked _____

| P5 | P6 | P7 | REPRODUCTION | | | | R5 | R6 | TYPING | | | | | LABELS | | | | Other | Specify |
|---------------------------|-------------------------------------|-----------------|-------------------------------------|-----------------------|--------------------------|-----------------------------|---------------------|------------------------|--------------------|--------------------|---------------------------|--------------------|---------------|-------------------------------|-------------------|----|----|-------------|---------|
| | | | R1 | R2 | R3 | R4 | | | T1 | T2 | T3 | T4 | T5 | L1 | L2 | L3 | L4 | | |
| Supervise & train P1 - P6 | Xerox maintenance, load clean, etc. | Xerox operation | Flexowriter maintenance, load, etc. | Flexowriter operation | Separate & arrange cards | Supervise and train R1 - R5 | Typing entire cards | Typing onto unit cards | Revision of typing | Typing Corrections | Supervise & train T1 - T4 | Spine label typing | Pocket typing | Spine label & pocket revision | Supervise L1 - L3 | | | 01 | 02 |
| | | | | | | | | | | | | | | | | | | Hour | |
| | | | | | | | | | | | | | | | | | | 8 - 9 | |
| | | | | | | | | | | | | | | | | | | 9 - 10 | |
| | | | | | | | | | | | | | | | | | | 10 - 11 | |
| | | | | | | | | | | | | | | | | | | 11 - 12 | |
| | | | | | | | | | | | | | | | | | | 1 - 2 | |
| | | | | | | | | | | | | | | | | | | 2 - 3 | |
| | | | | | | | | | | | | | | | | | | 3 - 4 | |
| | | | | | | | | | | | | | | | | | | 4 - 5 | |
| | | | | | | | | | | | | | | | | | | other hours | |
| | | | | | | | | | | | | | | | | | | Total | |

LS 1 December 1967

APPENDIX C

TECHNICAL MEMORANDUM NO. 261

TO: NEBHE File

FROM: William R. Nugent

SUBJECT: An Algorithm for Splitting the LC Call Number

DATE: February 8, 1968

1. INTRODUCTION

Man has been able to successfully split the atom and the infinitive, but the LC Call Number still defies simple solutions. A new algorithm is presented here. It works on Ann Curran's list of 39 examples (TM 233-9/26/67) and on LC's list of 6. These examples and the results of applying the algorithm to them are contained in pages C8 - C13 of this memo. The algorithm is a multiple pass process that first identifies kinds of elements in the linear string form of the call number, and then formats these according to their identity and the maximum column length. Sequences of alpha characters, numeric characters, and punctuation are used as a guide to identification and division. Beginning and end of line punctuation is generally dropped where item division is its only function, because end-of-line is itself a strong division, such punctuation is therefore redundant in column form, and conservation of space is achieved.

The Curran and LC examples are shown to illustrate the algorithm; interested staff members are invited to try it on other examples and determine what circumstances, if any, would cause the algorithm to incorrectly split.

2. ALGORITHM

The linear string of input symbols is taken in strict sequence from left to right and up to 3 classes of items are identified and extracted. The extracted items are then examined and divided where necessary to maintain a 6 character column. The items are then further scanned to concatenate certain short items occupying separate lines. The process, therefore, has 4 parts:

1. identification
2. extraction
3. division
4. concatenation

2.1 Identification and Extraction

Eight classes of items are identified, which we will refer to by their class numbers:

1. Alphabetic portion of class number
2. Initial numeric portion of class number
3. Decimal fraction portion of class number
4. Book number (initial alpha numeric portion)
5. Year designation
6. Alpha numeric designator (part numbers, etc.)
7. Other
8. Initial alphabetic designator (e.g., microfiche, incunabula, etc.)

Identification is performed on the basis of matching the first and successive characters of the input with a specific form of symbol string. The input string is identified successively left to right, and as one identification is made, the portion identified is extracted and tagged, and the identifier then works on the first and successive characters remaining in the input string. Tests for certain classes may be repeated twice, but they are not returned to later. Figure 1 shows the flow chart of the test sequence. The particular tests are described here subsequently.

The identification method seeks a predictable order of items, and when it correctly identifies these items, as it generally does, the resulting item partitioning is itself usually sufficient for determining column content. When identification is correct and the item is overly long, the identification helps in determining its final format. When a first unruly item is detected, the whole call number is assumed unruly and this and successive items are placed in class 7 ("other"), and further partitioning may be necessary.

The following definitions are used, defined via Backus conventions. Non-primitives are enclosed in triangular brackets, the vertical bar means "or" and the overbar means "not".

$$\langle \text{character} \rangle ::= \langle \text{alpha} \rangle \mid \langle \text{numeric} \rangle \mid \langle \text{punctuation} \rangle \mid \langle \text{space} \rangle$$

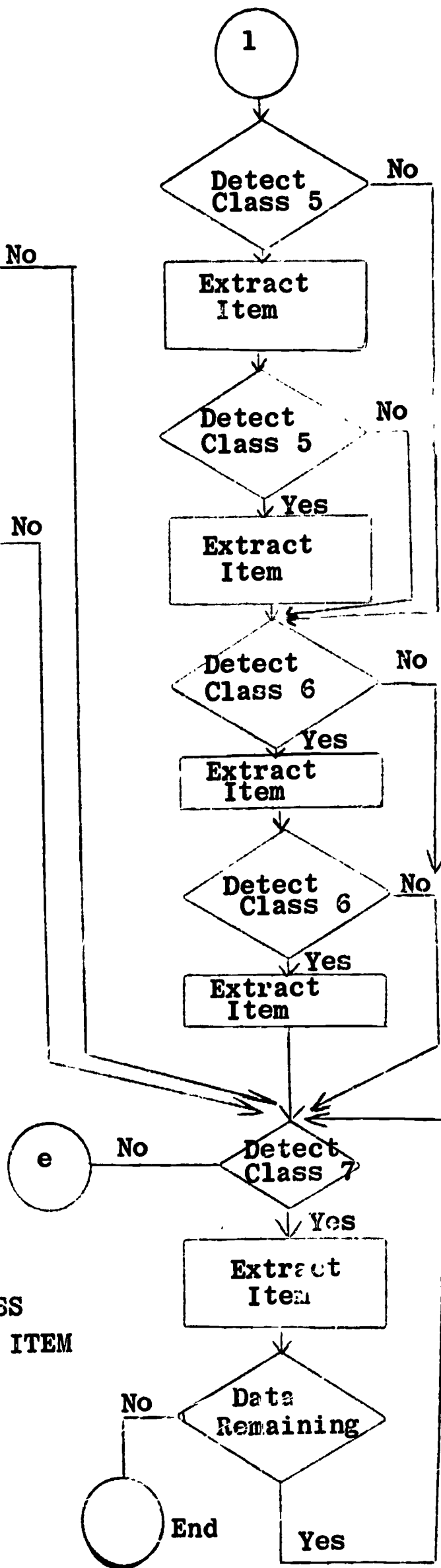
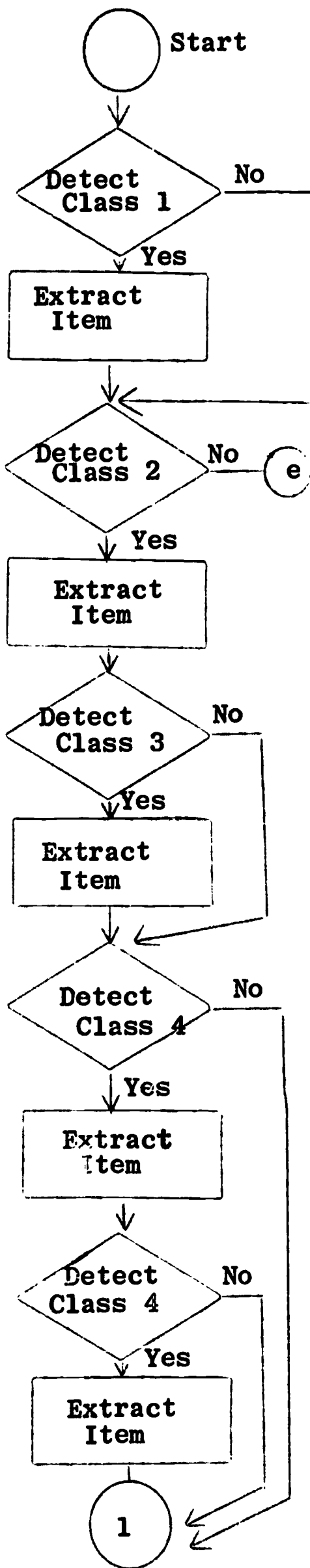
$$\langle \text{alpha} \rangle : = \langle \alpha \rangle : = A \mid B \mid C \mid \dots \mid Z \mid a \mid b \mid c \mid \dots \mid z$$

$$\langle \text{numeric} \rangle : = \langle \mathcal{N} \rangle : = 0 \mid 1 \mid 2 \mid \dots \mid 9$$

$$\langle \text{space} \rangle : = \langle \square \rangle$$

$$\langle \text{punctuation} \rangle : = \langle p \rangle : = \langle \text{slash} \rangle \mid \langle \text{comma} \rangle \mid \langle \text{period} \rangle \mid \langle \text{hyphen} \rangle \mid \langle \text{other code} \rangle$$

$$\langle \text{end of call number} \rangle : = \langle e \rangle : = \text{as determined by format and context}$$



SEQUENCE OF CLASS IDENTIFICATION AND ITEM EXTRACTION

FIGURE 1

<slash> : = /

<comma> : = ,

<period> : = .

<other code> : = any other character

To define the identification criteria and specify the symbols to be extracted, we will require sequence operators. We use conventions similar to Kleene's, where each symbol is a defining sequence, or parenthetical expression in a defining sequence, represents a discrete and necessary occurrence. The following conventions apply:

$\alpha\beta$: = initial alpha immediately followed by beta

I : = Identical character (universal)

$\alpha I \beta$: = initial alpha immediately followed by another character immediately followed by beta.

$\alpha \beta^J \alpha$: = initial alpha immediately followed by J betas immediately followed by alpha.

$\alpha \beta^* \alpha$: = initial alpha immediately followed by zero or more successive betas immediately followed by alpha, e.g., $\alpha\alpha$ or $\alpha\beta\alpha$ or $\alpha\beta\beta\alpha$, etc.

$\alpha^* \beta$: = $\beta \mid \alpha\beta \mid \alpha\alpha\beta \mid \dots$ etc.

$\alpha \bar{\beta}$: = initial alpha immediately followed by a character not beta

$\alpha(\beta/\gamma)$: = initial alpha immediately followed by a character either beta or gamma.

ϵ : = null character

$(\alpha/\epsilon) \beta \gamma$: = $\alpha\beta\gamma \mid \beta\gamma$: = initial character may be ϵ or β , followed respectively by $\beta\gamma$ and γ .

The identifying sequences, and extracted sequences are shown in Figure 2. In operation, the first character of the input string and its following symbols (or the first character of the input string after the last extracted item) must exactly match the class specification sought in the program flow of Figure 1. When a match is obtained, a portion of the matched data is extracted, as specified in the extraction list (Figure 2). Progressing through the flow of Figure 1, the next sought-for class

| <u>Class</u> | <u>Identity Starting with First Remaining Character</u> | <u>Characters Extracted from Identified Portion</u> |
|--------------|---|---|
| 1 | $\alpha\alpha^*n$ | $\alpha\alpha^*$ |
| 2 | $nn^*(\bar{n} e)$ | nn^* |
| 3 | $\cdot nn^*(\cdot \alpha e)$ | nn^* |
| 4 | $(\cdot \psi)\alpha\alpha^*n^*(\bar{n} e)$ | $\alpha\alpha^*n^*$ |
| 5 | $p^*\square n^4$ | n^4 |
| 5a | $(\alpha \alpha^2)(\square e)$ | $\alpha\alpha^*$ |
| 6 | $p^*\alpha\alpha^*(p p\square \psi)nn^*(\bar{n} e)$ | $\alpha\alpha^*(p p\square \psi)nn^*(- \psi)$ |
| 7 | $p^*\bar{p}I^*(\square e)$ | $\bar{p}I^*$ |
| 8 | $\alpha\alpha^*(p \psi)\square$ | $\alpha\alpha^*$ |

Item Classes, Identifying Sequences, and Characters Extracted.

FIGURE 2

specification is matched against the input character string starting with the character after the last extracted character. What we are doing, in brief, is identifying an item in terms of its possible forms and necessary pre-and post-delimiters, and then extracting the item only.

Item 6 will serve as an illustration for the rest, which are simpler. We identify a class 6 item by a sequence consisting of zero or more punctuation characters; one or more alpha characters, either punctuation, or punctuation space, or if neither of these exist then the character specification is null and the considered input symbol must match the next specification symbol; one or more numeric characters; an end of record or character not numeric. When this specified sequence is found, the extracted portion of it consists of the sequence less initial punctuation, if any, and through the final numeric portion, including at the end a hyphen if such should exist as the terminating \bar{n} character.

The remaining 7 items are considerably simpler. As each item is identified and extracted, it is put in a list and identified with a class label. Successive items of the same class are kept discrete, and appropriately labelled. The item lists resulting are illustrated for the examples in pages C3 - C13 of this memo.

2.2 Division

Items in the list are then examined for length, and where items exceed 6 characters they are divided into 2 items of the same class. Rules for division are somewhat class-dependent, but use subsets of the same rules in the same order.

2.2.1 Rules

- Rule 1: Start with 7th character and work back to 1st character, and break on first encountered space. If none, repeat and look for slash, if none repeat and look for comma. When a break point is found eliminate the space, slash, or comma from the item, and carry the remainder of the item forward as a new item of the same class.
- Rule 2: Check if item begins α^7 , if so replace 6th symbol with a hyphen and make new item from symbol 6 onward.
- Rule 3: Check if item begins $\alpha\alpha^*.nn^*\bar{n}$ and if so check if portion $\alpha\alpha^*.nn^*$ is 6 characters, if so, make a new item starting with \bar{n} . If portion $\alpha\alpha^*.nn^*$ is less than 6 characters, check if \bar{n} : = hyphen and begin new item with character after \bar{n} if so, or with \bar{n} if not so.

C - 7

Rule 4: Check from 7th character back for p symbol not period, if found, break, dropping the p symbol. If not found repeat and look for period, if found, break dropping the period unless it was preceded by an \ character.

Rule 5: Break after 6th character.

The rules are applied in the above order when an item is to be split. Rule 3 is only used in splitting items of classes 4, 6, 7 and 8. Rule 2 is only used for classes 7 and 8. However, no harm is done by applying these rules without exception to items of all classes - it is merely a waste of time. The exigencies of program construction can determine the preferred method.

2.3 Concatenation

After identification and division have been performed, one or two very short items may exist that are properly combined. Application of the following rules provides this:

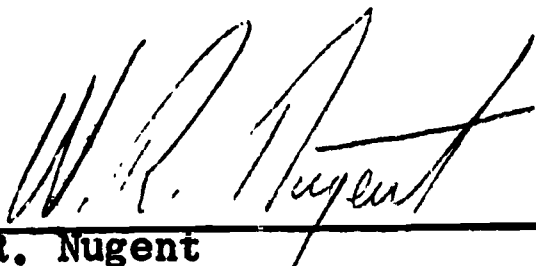
Rule 6: Check if successive items of class 2 and class 3 both exist, if so, add a period after the class 2 item add concatenate the class 3 item if the total is 6 characters or less.

Rule 7: Check for class 7 item with single character, concatenate into preceeding class 4 or class 7 item, if such exists and total character length does not exceed 6.

Rule 8: Check contiguous class 7 - class 7 pairs, and concatenate with intervening space if total does not exceed 6 characters.

3. CONCLUSIONS

The algorithm presented is a multiple pass process and the same rules could probably be incorporated into a single pass system, at no loss in accuracy and no necessary gain in simplicity. The algorithm does not result in a simple program. It works on 45 difficult examples. Simpler solutions should be tested on the same examples. Additional difficult examples should be tested with this algorithm.


W. R. Nugent

tmd

Examples of Algorithmic Call Number Division

1.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|-----|---------------|
| QL678.A45 | | |
| 1 | QL | same |
| 2 | 678 | |
| 4 | A45 | |

2.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|------|---------------|
| HF5548.2.A72 | | |
| 1 | HF | HF |
| 2 | 5548 | 5548.2 |
| 3 | 2 | A72 |
| 4 | A72 | |

3.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|------|---------------|
| HF5548.276A72 | | |
| 1 | HF | HF |
| 2 | 5548 | 5548 |
| 3 | 276 | .276 |
| 4 | A72 | A72 |

4.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|------|---------------|
| Z7164.E2A57 | | |
| 1 | Z | same |
| 2 | 7164 | |
| 4 | E2 | |
| 4 | A57 | |

5.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|-----|---------------|
| E467.1.A2A15 | | |
| 1 | E | E |
| 2 | 467 | 467.1 |
| 3 | 1 | A2 |
| 4 | A2 | A15 |
| 4 | A15 | |

6.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|------|---------------|
| HF1456 1964.C6 | | |
| 1 | HF | same |
| 2 | 1456 | |
| 5 | 1964 | |
| 6 | C6 | |

7.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|---------|---------------|
| BS191.A1.1952.N4 | | |
| 1 | BS | BS |
| 2 | 191 | 191 |
| 4 | A1 | A1 |
| 7 | 1952.N4 | 1952 |
| | | N4 |

8.

| <u>Class Items</u> | | <u>Format</u> |
|------------------------|-------|---------------|
| JX636 1892 1955, no.21 | | |
| 1 | JX | same |
| 2 | 636 | |
| 5 | 1892 | |
| 5 | 1955 | |
| 6 | no.21 | |

9.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|------|---------------|
| GA408 1751.F7M3 | | |
| 1 | GA | same |
| 2 | 408 | |
| 5 | 1751 | |
| 6 | F7 | |
| 6 | M3 | |

10.

| <u>Class Items</u> | | <u>Format</u> |
|--------------------|------|---------------|
| RD31.A425 1965 | | |
| 1 | RD | same |
| 2 | 31 | |
| 4 | A425 | |
| 5 | 1965 | |

11.

QK647.A35 1964c

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | QK | |
| 2 | 647 | same |
| 4 | A35 | |
| 5 | 1964c | |

12.

JX1975.A49 1923hd

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | JX | |
| 2 | 1975 | same |
| 4 | A49 | |
| 5 | 1923hd | |

13.

J87.M4 1958g no.2988

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | J | J |
| 2 | 87 | 87 |
| 4 | M4 | M4 |
| 5 | 1958g | 1958g |
| 6 | no.2988 | no. 2988 |

14.

QK1.U5 vol.35,pt.6

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | QK | |
| 2 | 1 | |
| 4 | U5 | same |
| 6 | vol.35 | |
| 6 | pt.6 | |

15.

Q11.P6 N.S.,vol.55pt.3

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|-----------------|------------------------|
| 1 | Q | Q |
| 2 | 11 | 11 |
| 4 | P6 | P6 |
| 7 | N.S.,vol.55pt.3 | N.S. vol.55 pt.3 |

16.

D1.T68 t.14,zesz.4

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | D | |
| 2 | 1 | |
| 4 | T68 | same |
| 6 | t.14 | |
| 6 | zesz.4 | |

17.

DA25.B5 [no.15]

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | DA | |
| 2 | 25 | same |
| 4 | B5 | |
| 6 | no.15 | |

18.

E51.H337 vol.57/no.1.

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | E | |
| 2 | 51 | |
| 4 | H377 | same |
| 6 | vol.57 | |
| 6 | no.1 | |

19.

Z733.C531J

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | Z | Z |
| 2 | 733 | 733 |
| 4 | C531 | C531J |
| 7 | J | |

20.

Z696.U5H 1959

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | Z | Z |
| 2 | 696 | 696 |
| 4 | U5 | U5H |
| 7 | H | 1959 |
| 7 | 1959 | |

21.

HJ9013.C6b

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | HJ | HJ |
| 2 | 9013 | 9013 |
| 4 | C6 | C6b |
| 7 | b | |

22.

FR2750.Bo7 1958

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PR | |
| 2 | 2750 | |
| 4 | Bo7 | same |
| 5 | 1958 | |

23.

PZ3.C3983Br

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | |
| 2 | 3 | |
| 4 | C3983 | |
| 7 | Br | |

24.

PZ3.C4637Cas2

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | |
| 2 | 3 | |
| 4 | C4637 | same |
| 6 | Cas2 | |

25.

PZ3.D89Th 80

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | PZ |
| 2 | 3 | 3 |
| 4 | D89 | D89 |
| 7 | Th | Th 80 |
| 7 | 30 | |

26.

PZ3.C86153Ghfc 2

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | PZ |
| 2 | 3 | 3 |
| 4 | C86153 | C86153 |
| 7 | Ghfc | Ghfc 2 |
| 7 | 2 | |

27.

PZ7.B4656Wav

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | |
| 2 | 7 | |
| 4 | B4656 | same |
| 7 | Wav | |

28.

PZ3.S38An

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | |
| 2 | 8 | |
| 4 | S38 | same |
| 7 | An | |

29.

PZ3.1.H75IWat

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | PZ |
| 2 | 8 | 8.1 |
| 3 | 1 | H75 |
| 4 | H75 | IWat |
| 7 | IWat | |

30.

PZ10.3.S837No

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | PZ |
| 2 | 10 | 10.3 |
| 3 | 3 | S837 |
| 4 | S837 | No |
| 7 | No | |

31.

PZ10.3.D632Lad4

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | PZ | PZ |
| 2 | 10 | 10.3 |
| 3 | 3 | D632 |
| 4 | D632 | Lad4 |
| 6 | Lad4 | |

32.

JX1977.A2 E/3111, etc.

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | JX | |
| 2 | 1977 | same |
| 4 | A2 | |
| 6 | E/3111 | |
| 7 | etc. | |

33.

JX1977.A2 E/CN.6/341/rev.1

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|------------------|------------------------|
| 1 | JX | JX |
| 2 | 1977 | 1977 |
| 4 | A2 | A2 |
| 7 | E/CN.6/341/rev.1 | E/CN.6 341 rev.1 |

34.

ML29c.M5

(music)

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | ML | |
| 2 | 29 | same |
| 7 | c.M5 | |

35.

G3701s.C5 var.U5

(map)

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | G | |
| 2 | 3701 | same |
| 7 | s.C5 | |
| 7 | var.U5 | |

36.

Microcard PS3545.Y45

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 0 | Microcard | Micro- |
| 1 | PS | card |
| 2 | 3545 | Ps |
| 4 | Y45 | 3545 Y45 |

37.

Microfilm 01104 no.93-94 AP

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|--------------------|
| 0 | Microfilm | Micro- |
| 7 | 01104 | card |
| 7 | no.93-94 | 01104 |
| 7 | AP | no.93- 94 AP |

32.

Microfilm 01291 reel 63, no.1E

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 0 | Microfilm | Micro- |
| 7 | 01291 | film |
| 7 | reel | 01291 |
| 7 | 63 | reel |
| 7 | no.1E | 63 no.1E |

39.

Incun. 1494.L3 Rosenwald Collec-
tion

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|-----------------------------------|
| 0 | Incun. | Incun. |
| 7 | 1494.L3 | 1494 |
| 7 | Rosenwald | L3 |
| 7 | Collection | Rosen- wald Collec- tion |

L.C.1

JX1977.A2 St/SOA/40

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | JX | JX |
| 2 | 1977 | 1977 |
| 4 | A2 | A2 |
| 7 | St/SOA/40 | St/SOA 40 |

L.C.2

UG633.A3763 no.62-5 1960

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | UG | |
| 2 | 633 | |
| 4 | A3763 | |
| 6 | no.62- | same |
| 7 | 5 | |
| 7 | 1960 | |

L.C.3

J85 1951. C9j 32d,1st no.17

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | J | J |
| 2 | 85 | 85 |
| 5 | 1951 | 1951 |
| 6 | C9 | C9j |
| 7 | j | 32d |
| 7 | 32d,1st | 1st |
| 7 | no.17 | no.17 |

L.C.4

DT333.S34 47.Jahrg.,Heft Nr.8

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|----------------|------------------------|
| 1 | DT | DT |
| 2 | 333 | 333 |
| 4 | S34 | S34 |
| 7 | 47.Jahrg.,Heft | 47 |
| 7 | Nr.8 | Jahrg. Heft Nr.8 |

L.C.5

Z2407.B no.32, etc.

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | Z | |
| 2 | 2407 | same |
| 7 | B | |
| 7 | no.32, | |
| 7 | etc. | |

L.C.6

BX2264.M56 1565 Rosenwald Coll.

| <u>Class</u> | <u>Items</u> | <u>Format</u> |
|--------------|--------------|---------------|
| 1 | BX | BX |
| 2 | 2264 | 2264 |
| 4 | M56 | M56 |
| 5 | 1565 | 1565 |
| 7 | Rosenwald | Rosen- |
| 7 | Coll. | wald Coll. |

Variations with respect to model examples in Ann Curran's TM-233

- | | |
|---|--|
| <p>(8) JX 636 1892 1955 no.21</p> | <p>example shows "1955," though final commas are missing in both example and algorithmic versions in the case of nos. 14 and 16 -- difference assumed trivial.</p> |
| <p>(17) DA 25 B5 no.15</p> | <p>example includes dropped indentures "[no.15]". -- difference assumed trivial.</p> |
| <p>(18) E 51 H337 vol.57 no.1</p> | <p>example has "vol.57/" slash conveys no information --difference assumed trivial.</p> |
| <p>(31) PZ 10.3 D632 Lad4</p> | <p>example has 'Lad 4', though input sequence has no space -- difference assumed trivial.</p> |

APPENDIX D

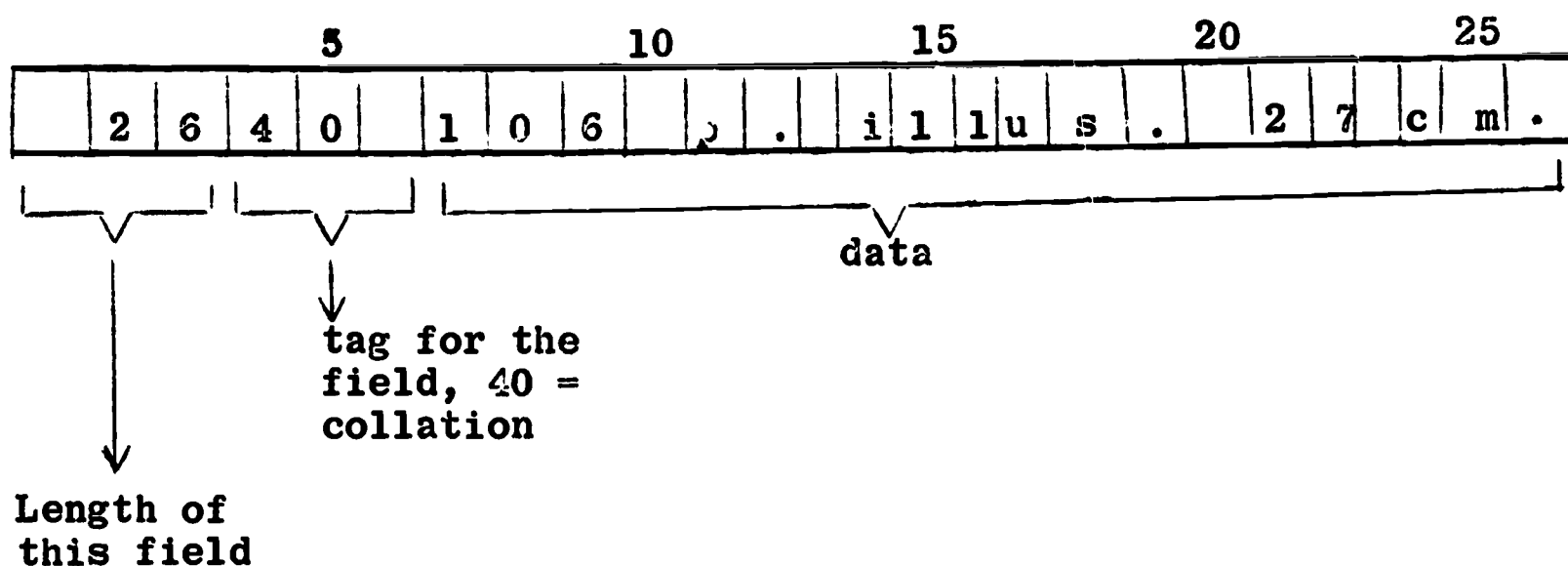
Analysis of the MARC II Format

The MARC II report defines format as "the structure, content, and coding of a record. The structure will provide the framework for incorporating both fixed and variable length fields within the record. The content is the data recorded in these fields. The coding is the machine representation of the character set."¹

Another significant characteristic of machine formats, in addition to the structure, content, and coding, is the identification of the data in the record. In this appendix the structure, content, and identification of data in the MARC I and MARC II formats will be compared. The implications that the changes in the MARC II format have on the NELINET processing of Library of Congress data and the generation of its own cataloging records will also be noted.

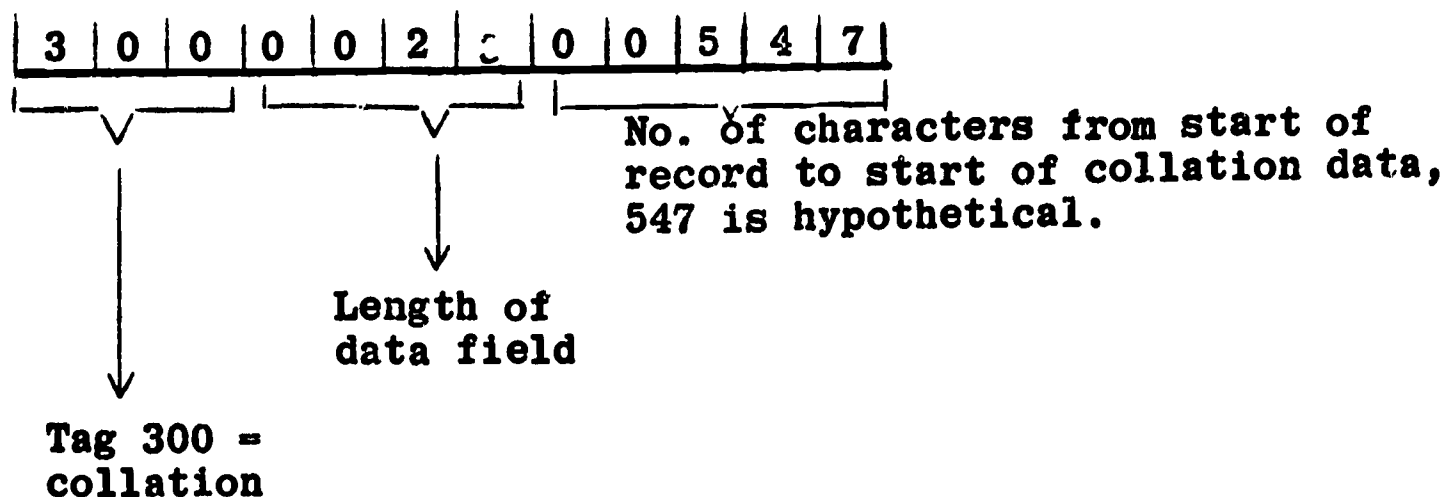
Structure

The MARC I format contains a fairly long fixed length field followed by a number of variable length fields. Each variable field contains a logical segment of the catalog card data preceded by six characters which contain the length of the field and the identifying tag for the field. The collation statement for a book would be found in the record as:

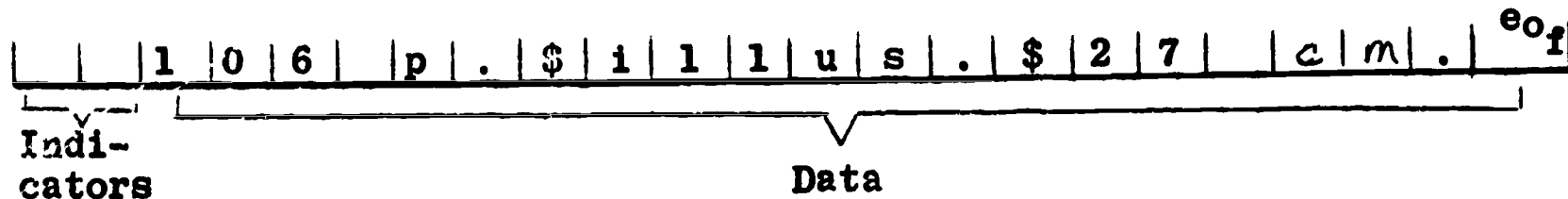


¹ Avram, Henriette D., Knapp, John F. and Rather, Lucia J., The MARC II Format. A Communications Format for Bibliographic Data. Preliminary Edition. Washington, Library of Congress, December, 1967.

The MARC II format consists of three parts - the leader, the directory, and the data fields. (See Figure D-1.) The leader is a short fixed length field which contains certain information about the machine record. The directory is made up of fixed length fields containing the identifying tag, the length of the corresponding data field, and its starting position relative to the beginning of the record. The data fields may be of fixed or variable length. The first two characters of each data field are allotted to indicators which further identify or describe the data field. In a MARC II record the collation data in the above example would appear in the directory as:



and in the data field as:



(\$-delimiter, a non printing separator of subfields)

(eof = end of field mark)

The NELINET format consists of two parts - the directory (map) and the data. Each entry in the directory is of fixed length and contains an identifying tag for a field (item) in the data and the position of the first character of the field relative to the beginning of the data. Data fields can be of fixed or variable length.

The structure of the MARC II format is more similar to NELINET than was MARC I. As far as structure is concerned, MARC II and NELINET are highly compatible, both being "mapped" records.

MARC II - Tape Structure

| Field 1 Leader | Field 2 Directory | Field 3 Variable Control Number | Field 4 Variable Fixed Fields | Field 5 Variable Data Field 1 | Field 6 Variable Data Field 2 |
|-------------------|----------------------|--|--|--|--|
|-------------------|----------------------|--|--|--|--|

Leader

| 1 | 5 | 6 | 7 | 8 | 9 | 12 |
|---|------------------|---------------|--------------|----------------|---|------------------------|
| | | | | | | |
| | Record Length | Record Status | Legend Count | Type of Record | | Bibliographic Level |

Directory

| 1 | 3 | 4 | 7 | 8 | 12 | 13 | 15 | 16 | 19 | 20 | 24 |
|---|-----|---|-----------------|---|-----------------------------------|----|-----|----|-----------------|----|-----------------------------------|
| | | | | | | | | | | | |
| | Tag | | Field Length | | Starting Character Position | | Tag | | Field Length | | Starting Character Position |

Variable Data Field

| 1 | 2 | | 1 | 2 | |
|-------------|-------------|------|-------------|-------------|------|
| Indicator 1 | Indicator 2 | | Indicator 1 | Indicator 2 | |
| | | Data | | | Data |

FIGURE D-1

Content

Very little MARC I data has been eliminated from the MARC II format. The codes for publisher and city of publication are not included but have been replaced by the publisher's prefix in the Standard Book Number and the country of publication code respectively. The code to indicate that a work is a grammar has been omitted.

A considerable amount of new data has been added to the MARC II record. Some of this new information is contained as codes and indicators in the variable fixed field, Tag 001, and some is contained as new variable data fields, e.g., the National Library of Medicine Call Number. An itemized listing of the MARC II fixed and variable fields along with an indication of the new fields not in MARC I are presented in Table D-1.

Identification of Data

In the MARC II format many of the data fields are more precisely identified and described than they were in MARC I. Most of these additional distinctions are made to facilitate machine filing according to the Library of Congress filing rules. The distinctions among personal, corporate, and conference name entries as well as those indicating filing treatment for personal and corporate name added entries greatly increase the number of different tags in the MARC II format. The fixed and variable fields which are identified differently in MARC II than they were in MARC I are shown in Table D-1.

Table D-1 lists the data elements identified by position in the fixed field and by a tag given to a variable field. In the MARC format data elements within a variable field are also identified. Usually a special character called a delimiter separates the subfields contained within a field. A personal name entry, for example, may contain four subfields - name, title, identifier, and relator. Unlike MARC I, MARC II indicates vacant subfields. The changes in internal identification within tagged fields are shown in Table D-2.

Implications for the NELINET Processing Center Programming

The changes in the MARC II format will mean almost total reprogramming of the card production programs. In addition to the new data in MARC II and the differences in identification of data there are other changes in the MARC II record which will require extensive changes in the programs. Title added entries are specified by an indicator in the title statement field rather than b, a separate tag as they were in MARC I. Series entries in which the author of the series is the main entry are to be generated by combining the main entry with the last part of the series statement rather than from data in one field. Such techniques are better suited to a

machine based system than were the practices of MARC I but will require considerable programming changes to accommodate them.

The decision has, therefore, been made to delay such reprogramming until MARC II has been firmly established and has been running for some time. This delay will also provide the opportunity of gaining more operating experience with the present programs to see how they might be improved. In order to provide continuous service after the Library of Congress changes to MARC II, a conversion program will be written which converts the Library of Congress MARC II data tapes to the NELINET MARC I format. These tapes can then be run through the existing card and label production programs and NELINET card production demonstration services need not be interrupted. The conversion from MARC II to NELINET master file format is a much simpler program to write than a new card production program and could be written before the Library of Congress starts to issue MARC II tapes in July, 1968.

Data Creation

The tagging of data for a MARC II record will be much more difficult than MARC I tagging. One would have less confidence in anyone but a trained cataloger doing or revising it. The distinctions made among personal and corporate name added entries to indicate filing treatment for instance, will require knowledge of the filing rules. Identifying personal names as forename or surname entries will at times be difficult especially when dealing with foreign names. Distinguishing geographic names from political jurisdictions will sometimes be difficult as will determining some period subdivisions. More time will be required to tag the record; more time will be required to train people to tag properly; more errors can be expected in the tagging; and more correction time will be required before "clean" records are obtained.

As originally planned, the six participating libraries would send original cataloging or Library of Congress non-MARC cataloging to the center and the center would tag and key it. Some of the tags in the MARC II format would be difficult to assign apart from the cataloging process. Such tags are:

- (1) The main entry is the publisher. (The publisher is omitted on the catalog card both when it is the main entry and when it is unknown.)
- (2) The language of the original and the language from which the work was translated. (This also occurred with MARC I.)
- (3) The book contains an index.

Since MARC II will become the standard format for the exchange of machine readable cataloging data among libraries, it is certainly desirable that others generating machine records use it. To do so will require considerable time and sophistication on the part of those who identify the items in the data. If the NELINET center is to generate complete MARC II records, it can only do so with help from the participating libraries.

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|------|------|--|
| | | | Leader |
| | | | Record Length |
| | | | Record Status, New |
| | | | Record Status, Changed |
| ** | | | Record Status, Deleted |
| | | | Record Status, Old |
| ** | | | Legend, Count |
| ** | | | Legend, Type, Language Materials, Printed (A) |
| ** | | | Legend, Type, Language Materials, Manuscript (B) |
| ** | | | Legend, Type, Music, Printed (C) |
| ** | | | Legend, Type, Music, Manuscript (D) |
| ** | | | Legend, Type, Maps, Printed (E) |
| ** | | | Legend, Type, Maps, Manuscript (F) |
| ** | | | Legend, Type, Motion Pictures and Filmstrips (G) |
| ** | | | Legend, Type, Microforms (H) |
| ** | | | Legend, Type, Sound-Recordings, Language (I) |
| ** | | | Legend, Type, Sound-Recordings, Music (J) |
| ** | | | Legend, Type, Pictures (K) |
| ** | | | Legend, Type, Computer Mediums, (L) |
| ** | | | Legend, Type, Authority Data-Names (X) |
| * | Data | | |
| ** | New | | |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|---|
| | | | Leader (continued) |
| ** | | | Legend, Type, Authority Data-Subjects (Y) |
| ** | | | Legend, Bibliographic Level, Analytical |
| | | | Legend, Bibliographic Level, Monographic |
| | | | Legend, Bibliographic Level, Serial |
| ** | | | Legend, Bibliographic Level, Collective |
| ** | | | Directory |
| | | | Tag |
| | | | Field Length |
| ** | | | Starting Character Position |
| | 000 | | Variable Control No. |
| | | | LC Card No. (Prefix, Year, Serial No.) |
| ** | | | LC Card No. Check Digit |
| | | | LC Card No. Supplement No. |
| | | | LC Card No. Suffix |
| | 001 | | Variable Fixed Field |
| ** | | | No. of Entries in Directory |
| ** | | | Date Entered on File (Day, Mo., Yr.) |
| | | | Type of Publication Date, Single (S) |
| ** | | | Type of Publication Date, Copyright (C) |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|---|
| | | | Variable Fixed Field (continued) |
| | | | Type of Publication Date, Not Known (N) |
| | | | Type of Publication Date, Reprint (R) |
| | | | Type of Publication Date, Multiple (M) |
| | | | Type of Publication Date, Questionable (Q) |
| | | | Date 1 |
| | | | Date 2 |
| ** | | | Country of Publication Code |
| | | | Illustration Code - Illustration (A) |
| | | | Illustration Code - Maps (B) |
| ** | | | Illustration Code - Portraits (C) |
| ** | | | Illustration Code - Charts (D) |
| ** | | | Illustration Code - Plans (E) |
| ** | | | Illustration Code - Plates (F) |
| ** | | | Illustration Code - Music (G) |
| ** | | | Illustration Code - Coats of Arms (I) |
| ** | | | Illustration Code - Genealogical Tables (J) |
| ** | | | Illustration Code - Forms (K) |
| | | | Intellectual Level Code (Juvenile) |
| ** | | | Form of Microreproduction Code, Microfilm (A) |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|---|
| | | | Variable Fixed Field (continued) |
| ** | | | Form of Microreproduction Code, Microfiche (B) |
| ** | | | Form of Microreproduction Code, Microopaque (C) |
| | | | Form of Content Code, Bibliographies (A) |
| ** | | | Form of Content Code, Catalogs (B) |
| ** | | | Form of Content Code, Indexes (C) |
| ** | | | Form of Content Code, Abstracts (D) |
| | | | Form of Content Code, Dictionaries (E) |
| ** | | | Form of Content Code, Encyclopedias (F) |
| ** | | | Form of Content Code, Directories (G) |
| ** | | | Form of Content Code, Yearbooks (H) |
| ** | | | Form of Content Code, Statistics (I) |
| ** | | | Form of Content Code, Handbooks (J) |
| ** | | | Government Publication Indicator |
| | | | Conference Publication Indicator |
| ** | | | Festschrift Indicator |
| ** | | | Index Indicator |
| * | | | Main Entry in Body of Entry Indicator |
| ** | | | Fiction Indicator |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|--------------------------------------|
| | | | Variable Fields |
| ** | 002 | | Legend Extension |
| | 003 | 0 | Language(s) - Non Translation |
| | 003 | 1 | Language(s) - Translation |
| | 010 | | LC Card No. |
| | 011 | | National Bibliography No. |
| ** | 012 | | Standard Book No. |
| ** | 013 | | PL 430 No. |
| ** | 014 | | Search Code |
| ** | 019 | | Local System No. |
| ** | 020 | | BNB Classification No. |
| | 030 | | Dewey Decimal Classification No.(s) |
| | 050 | 0 | LC Call No. |
| | 050 | 1 | LC Call No. - Book not in LC |
| | 051 | | Copy Statement |
| ** | 060 | | NLM Call No. |
| ** | 070 | | NAL Call No. |
| ** | 071 | | NAL Subject Category No. |
| ** | 080 | | Universal Decimal Classification No. |
| ** | 090 | | Local Call No. |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Elements |
|--------|-----|------|--|
| | | | Variables Fields (continued) |
| | | | Main Entry |
| * | 100 | 0 | Personal, Forename |
| * | 100 | 1 | Personal, Single Surname |
| * | 100 | 2 | Personal, Multiple Surname |
| * | 100 | 3 | Personal, Name of Family |
| ** | 100 | 4 | Personal, Forename, is Subject |
| ** | 100 | 5 | Personal Single Surname, is Subject |
| ** | 100 | 6 | Personal, Multiple Surname, is Subject |
| ** | 100 | 7 | Personal, Name of Family, is Subject |
| * | 110 | 0 | Corporate, Surname |
| * | 110 | 1 | Corporate, Place |
| * | 110 | 2 | Corporate, Name (Direct Order) |
| ** | 110 | 4 | Corporate, Surname, is Subject |
| ** | 110 | 5 | Corporate, Place, is Subject |
| ** | 110 | 6 | Corporate, Name (Direct Order), is Subject |
| * | 111 | 0 | Conference, Surname |
| * | 111 | 1 | Conference, Place |
| * | 111 | 2 | Conference, Name (Direct Order) |
| ** | 111 | 4 | Conference, Surname, is Subject |
| ** | 111 | 5 | Conference, Place, is Subject |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|---|
| | | | Variable Fields (continued) |
| | | | Main Entry (continued) |
| ** | 111 | 6 | Conference, Name (Direct Order), is Subject |
| * | 120 | 0 | Corporate with Form Subheading |
| ** | 120 | 1 | Corporate with Form Subheading, is Subject |
| | 130 | 0 | Uniform Title |
| ** | 130 | 1 | Uniform Title, is Subject |
| | 200 | 0 | Title ("Uniform," "Conventional," or "Filing") |
| ** | 200 | 1 | Title ("Uniform," "Conventional," or "Filing") is on LC Printed Cards |
| ** | 210 | 0 | Romanized Title |
| ** | 210 | 1 | Romanized Title, Receives Title Added Entry |
| ** | 220 | | Translated Title |
| | 240 | 0 | Title Statement |
| * | 240 | 1 | Title Statement Receives Added Entry |
| | 250 | | Edition Statement |
| | 260 | | Place(s) |
| | 261 | 0 | Publisher |
| ** | 261 | 1 | Main Entry is Publisher |
| | 262 | | Date(s) |
| | 300 | | Collation |
| | 350 | | Bibliographic Price |
| ** | 360 | | Converted Price |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Elements |
|--------|-----|------|--|
| | | | Variable Fields (continued) Series, Traced, |
| * | 400 | 0 | Personal, Forename |
| * | 400 | 1 | Personal, Single Surname |
| * | 400 | 2 | Personal, Multiple Surname |
| * | 400 | 3 | Personal, Name of Family |
| ** | 400 | 4 | Personal, Forename, is Main Entry |
| ** | 400 | 5 | Personal, Single Surname, is Main Entry |
| ** | 400 | 6 | Personal, Multiple Surname, is Main Entry |
| ** | 400 | 7 | Personal, Name of Family, is Main Entry |
| * | 408 | | Personal, Title |
| * | 410 | 0 | Corporate, Surname |
| * | 410 | 1 | Corporate, Place |
| * | 410 | 2 | Corporate, Name (Direct Order) |
| ** | 410 | 4 | Corporate, Surname, is Main Entry |
| ** | 410 | 5 | Corporate, Place, is Main Entry |
| ** | 410 | 6 | Corporate, Name (Direct Order), is Main Entry |
| * | 411 | 0 | Conference, Surname |
| * | 411 | 1 | Conference, Place |
| * | 411 | 2 | Conference, Name (Direct Order) |
| ** | 411 | 4 | Conference, Surname, is Main Entry |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Date Element |
|--------|-----|------|--|
| | | | Variable Fields (continued) |
| | | | Series, Traced, |
| ** | 411 | 5 | Conference, Place, is Main Entry |
| ** | 411 | 6 | Conference, Name (Direct Order), is Main Entry |
| * | 418 | | Corporate, Title |
| | 440 | | Title |
| | 490 | 0 | Series, Untraced |
| * | 490 | 1 | Series, Untraced, Series is Traced Differently |
| * | 500 | | Bibliography Note |
| * | 510 | | Dissertation Note |
| * | 520 | | Contents Note (Formatted) |
| * | 530 | | "Bound With" Note |
| * | 540 | | "Limited Use" Note |
| * | 550 | | General Notes (All Others) |
| ** | 560 | | Annotation or Abstract |
| | | | Subject Entry |
| * | 600 | 0 | Personal, Forename |
| * | 600 | 1 | Personal, Single Surname |
| * | 600 | 2 | Personal, Multiple Surname |
| * | 600 | 3 | Personal, Name of Family |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|---|
| | | | Variable Fields (continued) |
| | | | Subject Entry |
| * | 608 | | Personal, Title |
| * | 610 | 0 | Corporate, Surname |
| * | 610 | 1 | Corporate, Place |
| * | 610 | 2 | Corporate, Name (Direct Order) |
| * | 611 | 0 | Conference, Surname |
| * | 611 | 1 | Conference, Place |
| * | 611 | 2 | Conference, Name (Direct Order) |
| * | 618 | | Corporate, Title |
| * | 620 | | Corporate, Name with Form Subheading |
| * | 628 | | Corporate, Name with Form Subheading, Title |
| * | 630 | | Uniform Title |
| * | 650 | | Topical |
| * | 651 | | Geographic Name |
| * | 652 | | Political Jurisdiction |
| * | 653 | | Proper Names, Not Capable of Authorship |
| * | 655 | | General Subdivision |
| * | 656 | | Period Subdivision |
| * | 657 | | Place Subdivision |
| ** | 660 | | NLM (MeSH) Heading |
| ** | 661 | | NLM (MeSH) Heading, Topical Subheadings |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|--|
| | | | Variable Fields (continued) |
| | | | Subject Entry |
| ** | 662 | | NLM (MeSH) Heading, Geographical Subheadings |
| ** | 663 | | NLM (MeSH) Heading, Period Subheadings |
| ** | 664 | | NLM (MeSH) Heading, Form Subheadings |
| ** | 670 | | NAL Agricultural/Biological Vocabulary |
| ** | 690 | | Local Subject Heading |
| | | | Added Entry |
| * | 700 | @ | Personal, Forename, Alternative |
| * | 700 | A | Personal, Single Surname, Alternative |
| * | 700 | B | Personal, Multiple Surname, Alternative |
| * | 700 | C | Personal, Name of Family, Alternative |
| * | 700 | D | Personal, Forename, Secondary |
| * | 700 | E | Personal, Single Surname, Secondary |
| * | 700 | F | Personal, Multiple Surname, Secondary |
| * | 700 | G | Personal, Name of Family, Secondary |
| * | 700 | H | Personal, Forename, Analytical |
| * | 700 | I | Personal, Single Surname, Analytical |
| * | 700 | J | Personal, Multiple Surname, Analytical |
| * | 700 | K | Personal, Name of Family, Analytical |
| | 708 | | Personal, Title |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Elements |
|--------|-----|------|---|
| | | | Variable Fields (continued) Added Entry |
| * | 710 | @ | Corporate, Surname, Alternative |
| * | 710 | A | Corporate, Place, Alternative |
| * | 710 | B | Corporate, Name (Direct Order), Alternative |
| * | 710 | D | Corporate, Surname, Secondary |
| * | 710 | E | Corporate, Place, Secondary |
| * | 710 | F | Corporate, Name (Direct Order), Secondary |
| * | 710 | H | Corporate, Surname, Analytical |
| * | 710 | I | Corporate, Place, Analytical |
| * | 710 | J | Corporate, Name, (Direct Order), Analytical |
| * | 711 | @ | Conference, Surname, Alternative |
| * | 711 | A | Conference, Place, Alternative |
| * | 711 | B | Conference, Name, (Direct Order), Alternative |
| * | 711 | D | Conference, Surname, Secondary |
| * | 711 | E | Conference, Place, Secondary |
| * | 711 | F | Conference, Name, (Direct Order), Secondary |
| * | 711 | H | Conference, Surname, Analytical |
| * | 711 | I | Conference, Place, Analytical |
| * | 711 | J | Conference, Name, (Direct Order), Analytical |
| | 713 | | Corporate, Title |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag | Ind. | Data Element |
|--------|-----|------|--|
| | | | Variable Fields (continued) |
| | | | Added Entry |
| * | 720 | 0 | Corporate, with Form Subheading, Alternative |
| * | 720 | 1 | Corporate, with Form Subheading, Secondary |
| * | 720 | 2 | Corporate, with Form Subheading, Analytical |
| * | 723 | | Corporate, with Form Subheading, Title |
| * | 730 | 0 | Uniform Title, Alternative |
| * | 730 | 1 | Uniform Title, Secondary |
| * | 730 | 2 | Uniform Title, Analytical |
| | 740 | | Title Traced Differently |
| * | 753 | | Proper Names Not Capable of Authorship |
| | | | Series Added Entry |
| * | 800 | 0 | Personal, Forename |
| * | 800 | 1 | Personal, Single Surname |
| * | 800 | 2 | Personal, Multiple Surname |
| * | 800 | 3 | Personal, Name of Family |
| * | 808 | | Personal, Title |
| * | 810 | 0 | Corporate, Surname |
| * | 810 | 1 | Corporate, Place |

TABLE D-1

MARC II - Identification of Fields

| Status | Tag. | Ind. | Data Elements |
|--------|------|------|--|
| | | | Variable Fields (continued) Series, Added Entry |
| * | 810 | 2 | Corporate, Name (Direct Order) |
| * | 811 | 0 | Conference, Surname |
| * | 811 | 1 | Conference, Place |
| * | 811 | 2 | Conference, Name (Direct Order) |
| * | 818 | | Corporate, Title |
| | 840 | | Title |

TABLE D-2

MARC II - Internal Identification Within Tagged Fields

| Status | Tag(s) | Data Element |
|--------|-------------|--|
| | 003 | Language(s) of Work |
| ** | 003 | Language Translated From |
| | 003 | Language of Original |
| ** | 003 | Language(s) of Summaries |
| ** | 012 | Standard Book Number--Publisher's Prefix |
| ** | 012 | Standard Book Number--Title No. |
| ** | 012 | Standard Book Number--Check Digit |
| ** | 030 | More than 1 Dewey Decimal Classification No. |
| * | 050 | LC Call No.--Class No. |
| * | 050 | LC Call No.--Book No. |
| * | 051 | Copy Statement--Class No. |
| * | 051 | Copy Statement--Book No. |
| * | 051 | Copy Statement--Remainder |
| ** | 060 | NLM Call No.--Class No. |
| ** | 060 | NLM Call No.--Book No. |
| ** | 090 | Local Call No. |
| ** | 090 | Local Holding Collection Code |
| ** | 090 | Local Number of Copies |
| | 100,400,600 | Personal Name Entries--Name |
| | 700,800 | Personal Name Entries--Title |
| | | Personal Name Entries--Identifier |
| | | Personal Name Entries--Relator |

* Data Identified Differently

** New Data

TABLE D-2

MARC II - Internal Identification Within Tagged Fields

| Status | Tag(s) | Data Element |
|--------|-------------|--|
| * | 110,410,610 | Corporate Entries { Name Subdivision <u>or</u> { Place Name Subdivision |
| * | 710,810 | |
| * | | |
| * | 111,411,611 | Conference Entries--Name |
| * | 711,811 | Conference Entries--No. |
| * | | Conference Entries--Place |
| * | | Conference Entries--Date |
| * | 120,620,720 | Corporate Name With Form Subheadings--Name |
| * | | Corporate Name With Form Subheadings--Title |
| | 240 | Title Statement--Short Title |
| | 240 | Title Statement--Remainder of Title |
| | 240 | Title Statement--Remainder to Edition |
| | 250 | Edition Statement--Edition Information |
| | 250 | Edition Statement--Remainder |
| * | 300 | Collation--Pagination |
| * | 300 | Collation--Illustrations |
| * | 300 | Collation--Height |
| ** | 300 | Collation--Thickness |
| * | 408,418,448 | Series--Title |
| * | 808,818,840 | Series--Number |

APPENDIX E

PRODUCTION PROBLEMS AND DECISIONS

Soon after New Hampshire began receiving cards and labels they requested an opportunity to meet with Inforonics about problems they had encountered. On February 8, 1968, Inforonics personnel went to the NEBHE office in Durham to discuss the output products with the UNH catalogers. Listed are the problems, questions, etc. considered and the decisions on them reached during the meeting or subsequent to it. In some cases, a decision was postponed until the other participants have received and considered some cataloging products.

1. Typeface of labels: Other typefaces are being investigated.
2. Spacing:
 - a. Before edition statement: No NEBHE standard, NELINET cards follow format of UConn, UMe, URI, and UMass, 2 spaces.
 - b. Before imprint: No NEBHE standard, NELINET cards follow format of UConn, UMe, UMass, and UVt, 2 spaces.
 - c. Within collation: Items are not separately identified on MARC tapes so no change is now feasible. NELINET libraries vary.
 - d. Before series statement: No NEBHE standard, three libraries use 3 spaces and three use 2 spaces. NELINET cards will use 2, as, other things being equal, the saving of space seems desirable.
 - e. Between tracing number and tracing data: No NEBHE standard, UConn, UMe, UMass and UVt leave one space, UNH leaves none. NELINET cards will have one space for the present. This space could be omitted if the libraries thought the space saved warranted it.
3. Blank lines:
 - a. Before the first note: All libraries follow this practice. NELINET cards do not in order to conserve space. This could be changed.
 - b. Before the tracings: Five libraries follow this practice (URI format unknown). NELINET cards do not in order to conserve space. This could be changed.
 - c. Between the header data on card 2 and continuation of text: This blank line is now produced on the NELINET cards.
4. The BNB number which can be used to order BNB cards will no longer be printed on NELINET CARDS; in its place an abbreviation

of the requesting library will be printed to aid card distribution.

5. Oversize books: Libraries may encounter oblong books whose width requires they be shelved as an oversize volume though their spine length does not. Present programs will probably not handle them correctly; libraries should submit the local call number with their oversize symbol.
6. Errors in data such as Robart for Robert are due to mistakes in the L.C. tapes.
7. Partial title entries: Programs are being corrected to produce "catchword" titles.
8. Series tracings: Programs that produce overprinted series headings are being corrected. When local series entry practices differ from L.C. treatment, the library will have to adjust the cards or its practice.
9. Erratic indentation of headings on card two's has been fixed.
10. Branch and special shelf location will appear over all call numbers for the present. This was agreed to by the six libraries at the meeting on May 24, 1967.
11. Author/title or title information repeated on card two's will be in L.C. format.
12. To reduce the frequency of card two's, line 14 will be made available to data and the continuation statement (Cont. next card) will be on the 15th line.
13. Tracings beginning on top line of the card: This is done to allow the maximum space for the tracings and is L.C. MARC practice. One could even assert it assists the catalog user to distinguish the heading data from the main entry.
14. Punctuation before a subject subdivision: MARC uses a double dash as do UConn and UVt. (UMass and UNH use space dash space and UMe uses a single space). To do other than follow MARC would require considerable effort, require more space and conform to the format of the same number of libraries.
15. Subject subdivision abbreviations: These are now spelled out.
16. It was suggested that in tracings "Title." and "Title:" be shortened to "T." and "T:" to save space. No NELINET libraries presently do so. The change, while fairly simple, seems best postponed until all the libraries have received some cards.

17. New Hampshire felt that inserting the cards and labels into envelopes was unnecessary. It will be suspended. If other libraries request it, resumption could be reconsidered.
18. Margins:
 - a. Labels: a one space left margin will be used on labels.
 - b. Cards:
 1. A one space margin to the left of the call number will be used.
 2. Programs presently produce cards with L.C. format, that is, the main entry begins to the left of all other lines. This means considerable wasted space as all subsequent lines begin at the second or third indentation. Changes to bring cards closer to the formats of the NELINET libraries are being written. "Hanging indention" will be used for the title paragraph of anonymous works as on p. 65 of L.C. Colvin's Cataloging sampler. (cf. Cataloging Service bulletin no. 69 and AA rules footnote p. 192). Line endings will not be indicated for titles in verse, etc. as the MARC record does not provide for them (cf. AA rule 133).
19. L.C. call number: Present programs to divide the L.C. call number for labels and for the left margin (gutter) of the cards are imperfect.
 - a. When the numeric part of the classification number is less than six characters long but includes a decimal point, the programs divide on the decimal which is contrary to NEBHE library practice and to the format specified in the Pre-sentation notes for the May 24, 1967 NEBHE meeting.
 - b. When a period is followed by a space, e.g., vol. 2, two carriage returns are executed resulting in a blank line between "vol." and "2".
 - c. The decimal point is not being dropped before the book number as was also specified in the May 24, 1967 presentation notes.
 - d. An algorithm to solve the problems of L.C. call number division has been developed by W. Nugent. Programming to incorporate it into the present programs has been postponed to allow manual testing of it. The present effect of the above problems should be small because "a" and "b" occur infrequently and "c" is not a problem until Rhode Island or Massachusetts begins requesting cards because New Hampshire presently retains the decimal point before the book number.

20. Title tracing: Present programs remove punctuation other than periods, question marks, and exclamation points from titles before using them as a heading. New Hampshire expressed the desire that a period appear at the end of all such headings. This programming is possible but has been postponed until other libraries have a chance to react to the cards.
21. Future problems: To enable Inforonics to assess and improve the quality of the products, New Hampshire will fill out problem reports for each new problem and annotate and return cards which are not usable. Both will be forwarded to R. Simmons. As soon as 7, 8, and 12 are solved, New Hampshire will resume requests.

APPENDIX F

OPERATING PROCEDURES FOR INFORMATION TRANSMITTAL IN THE NEW ENGLAND LIBRARY INFORMATION NETWORK

Prepared by

Robert H Simmons

October 18, 1967

OPERATING PROCEDURES FOR
INFORMATION TRANSMITTAL IN
THE NEW ENGLAND LIBRARY INFORMATION NETWORK

Installed at each of the participating libraries is a teletypewriter, model ASR-33. These teletypewriters will be used to transmit requests between the participating libraries and the NELINET central processing center. The requests will be for catalog cards, book spine labels, and book pocket labels, for the participating libraries and their branches.

The requesting procedures have been made as concise as possible to facilitate ease of use by the cataloger, and ease of transmittal by the transmitting clerk. Suggestions for easier methods are actively solicited.

NELINET Request Form

The NELINET Request Form is similar for each of the participating libraries. The area of deviation is the library code and the shelf list and branch name abbreviations. The request form will eventually have more information included in it as the demonstration period progresses into title and/or author searching.

A short explanation of each line follows. A more comprehensive explanation of each line is included in the section entitled "Cataloger's Procedures for Filling Out Request Sheets". We will use Figure 1, page 12 as an example work sheet.

- RI This will be preprinted on all Rhode Island library request sheets. This is a library identifier and will be transmitted with every request.
- LC This line is for Library of Congress call numbers. All requests for information will have this line filled in. At a later stage in the project this may not be known and appropriate author and title lines will be added.
- AT The next four lines in the sheet allows the user to specify where the books are destined to be sent, the number of copies, whether they are going onto a special shelf, and whether multivolume sets are involved.
- CN This line is available for libraries who wish to have a local call number placed on their cards and labels instead of the Library of Congress call number. This allows those participating libraries who use Dewey Decimal call numbers to use the system. This does not affect the normal procedure of getting the Library of Congress call number on the bottom of the catalog cards.

"Xtra Cards" If the library wishes to have extra main entry cards produced above the normal set, then by writing in the amount this will be accomplished.

"No Cards, Labels, Pockets" This is a multi-use line in that the library may suppress the production of catalog cards, spine labels, and book pockets, or any combination of the three outputs.

Cataloger Procedure for Filling Out the
NELINET Request Form

Whenever a cataloger wishes to request catalog cards, book spine labels, and/or book pocket labels through the New England Library Information Network Processing Center he will have to complete a NELINET Request Form. This form has been made as short as possible and, we hope, as easy as possible to fill out. Part of the data has been preprinted on the forms to further facilitate fast and easy completion by the cataloger. A separate form has been created for each participating library differing only in the library code name and the listing of abbreviations of shelf list names and branch library names.

The request forms are to be filled out in the following

ay:

Line Name

O, MA, ME,
H, RI, or

T

This line is the library identifier. One of the previous state abbreviations will be preprinted on each of the request forms. No action needs to be taken by the cataloger as this will be transmitted by the teletype operator.

C

The cataloger must enter the Library of Congress card number on this line. The cataloger is to write this number in the form in which it normally appears on L.C. cards.

Example:

64-4302
HEW 66-2362

T

This line, plus the next three lines, are provided for the cataloger to enter the location (copy is at....), copy-shelf statement and volume information. Some examples of how this information is to be filled out are given in pages 4 - 9.

CN

If the cataloger wishes to have a call number that is different from the Library of Congress call number then he may write this number on this line. This call number, referred to as a local call number, will then be placed in the left margin of all cards produced by this request. The local call number will also be used on all book spine labels, and book pocket labels produced by this request. If this line is not used, then the Library of Congress call number that appears on the bottom of the L.C. card will be automatically placed in the left margin of the catalog cards, and all labels produced by this request will contain the L.C. call number. Refer to page 10 for the proper way to write this number on the cards.

XC

If the cataloger wishes to have more than the normal amount of main entry cards (Xtra Cards) produced by the system for this request, the cataloger will write the extra amount he wishes in this space. This amount of extra main entry cards will be produced for this request in addition to the normal output.

NCLP

This line will be used by the cataloger if he wishes to prevent the output of catalog cards, book spine labels, and/or book pocket labels, (No Cards, Labels, Pockets). The cataloger must circle the C if cards are to be suppressed, the L if book spine labels are to be suppressed and/or the P if book pocket labels are to be suppressed.

This appendix will explain the proper way to fill out the location - copy shelf location, and volume statement. This line, which we will refer to as the technical process statement, or TPS, is used to produce the proper number of labels and select the correct profile for catalog cards. The TPS is made up of three separate parts:

1. Location This may be the branch abbreviation or the main library.
2. Copy - shelf statement The copy information and shelf location information is inserted in this section.
3. Volume information If the book consists of more than one volume they are noted in this section.

The parts are filled out in the following fashion:

1. Location The cataloger is to insert in this space the branch abbreviation or the word "main".

Example 1:

A request is made for the central library.

Solution:

Write m a i n in the space marked "location".

Example 2:

A request is made for a branch library.

Solution:

Write the branch abbreviation in the space marked "location".

Example 3:

A request is made for both a branch location and the central library.

Solution:

Write m a i n on the first line,
and the branch abbreviation on the
second line.

A semicolon must be placed between location and copy information. This indicates the end of the location section and the beginning of the copy - shelf statement section.

2. Copy-shelf
statement

The cataloger is to insert in this space the copy information and the shelf abbreviation if applicable. The following examples are given as guides.

Example 1:

The cataloger requests cards for a single book and it is copy 1.

Solution:

The cataloger will write c.1 in the space provided.

Example 2:

The cataloger requests cards (and/or labels) for three books, copies 1, 2, and 3.

Solution: write |c.1-3

Meaning:

The " | " means to enumerate the information following the bar. The information supplied in this case meant the span covered was c.1, c.2, and c.3.

Example 3:

The cataloger requests labels for copies 1-4 of a book and copy 4 is to go to the reference section.

Solution: write |c.1-3, c.4 Ref

The bar (|) means to enumerate the following information (up to a comma) if more than one copy-shelf statement item is listed. The comma separates subparts of the copy-shelf statement c.4 Ref means the labels produced for copy 4 will have Ref printed on them.

Example 4:

The cataloger request labels for 5 books, 2 to go to reference, 1 to the central library and 2 to the Blatz collection.

Solution:

c.1, | c.2-3 Ref, | c.4-5 Blatz

If volume information is involved then a semicolon will be used to separate copy information from volume information.

3. Volume information

The cataloger is to insert volume information in this space if it is present. The following examples are given as guides.

Example 1:

A single volume is involved (assuming this is the first of several to be issued in a set).

Solution:

v.1

Example 2:

Two volumes, volumes 1 and 2 are involved.

Solution:

| v.1-2

As in the copy-shelf statement the bar (|) means to enumerate the information following, up to a comma, if one is present.

F - 7

Example 3:

Two volumes are involved volumes 2 and 4.

Solution:

v.2, v.4

The comma separates subsections of the volume information part of the Technical Processing Statement.

Example 4:

Four volumes are involved volumes 1, 2, 3, and 5.

Solution:

|v.1-3, v.5

The bar signifies that the information following is to be enumerated. The comma ends the part to be enumerated.

Assume that the library wished to request processing cards for a single book. The statement will be the following:

AT ☐ MAIN

Note: c.1 does not have to be inserted if only 1 copy is present and no volume numbers are involved.

Assume the library wishes to request processing for the second copy of the above book at a later time.

AT ☐ main; c.2

Assume the library wishes to request processing for two copies of a book 1 and 2.

AT ☐ main; | c.1-2

for five copies, 1, 2, 3, 4, and 5

AT ☐ main; | c.1-5

The present copy span can be up to 99 copies.

Assume the library wishes to request processing for four copies, a replacement for copy one, which was destroyed and copies four, five, and six.

AT ☐ main; c.1, |c.4-6

If the copy one above was going to a shelf location, say reference, then the AT statement would look like this:

AT ☐ main; c.1 Ref, |c.4-6

Assume copy one was for the central library, copy two to four for the Ref collection, and copy five for the xyz shelf collection.

AT ☐ main; c.1, |c.2-4 Ref, c.5 xyz

Assume that on the above order two volumes v.1 and v.2 are involved.

AT ☐ main; c.1, |c.2-4 Ref, c.5 xyz; |v.1-2

Assume that only volume one is going to the xyz shelf location.

AT ☐ main; c.1, |c.2-4 Ref, |v.1-2

and

AT ☐ main; c.5 xyz; v.1

Assume copy 1 to main library

copy 2 to mathematics branch

copy 3 to reference collection.

AT ☐ main; c.1, c.3 Ref

and

AT ☐ MATH; c.2

Probably the most common request will be a single copy for a single location which will look like:

AT ☐ main

The following is an explanation of each character in an example statement.

Example:

AT ☐ main; | c.1-3, c.4 Ref; | v.1-3

AT ☐ = the prefix of the statement which allows the program to interpret the statement as a technical processing statement.

main = the location for which the request is made. Up to six characters may be entered in this space. If any entry is made other than the word 'main' the line will then be considered information for a Branch library and this location will appear on all output.

; = the semicolon separates the location information from the copy-shelf information.

| = the bar indicates that information following up to a comma or semicolon must be enumerated.

c.1-3 = the spread of copies to be enumerated.

, = separates the sub-parts of the statement.

c.4 Ref = the copy number and its shelf location. This automatically will be placed on all the catalog cards on labels produced for this copy number, (or numbers if it is to be enumerated).

; = separates the copy-shelf statement from the volume information.

| = enumerate the following volume information up to a comma or end of the statement.

v.1-3 = the spread of volumes to be enumerated.

A local call number or CN field on the request sheet is available for the libraries who wish to have a call number other than the Library of Congress call number on the spine labels, and the margin of catalog cards. This provision allows those libraries who are using the Dewey Decimal Call Number system or their own unique call numbering system to use the NELINET.

Certain arbitrary conventions are used to indicate upper case characters, but these conventions will only concern the transmittal clerk rather than the catalogers. The following are examples of formatting for local call numbers in this field and how the number will appear on the output medium.

Example 1:

574/.92074

Will look like:

547
.92074

The slash indicates the end of a line (it is interpreted as a carriage return). No more than six characters including punctuation may appear on a single line.

Example 2:

QP/481/.H4813

Will look like:

QP
481
.H4813

Example 3:

QL/478/.M25/1964

Will look like:

QL
478
.M25
1964

Sample Request Sheets

The following is a number of sample request sheets that have been filled in with request data.

In Figure 1, the library is requesting cards, spine labels and pocket labels for a single copy for the main library. The Library of Congress call number is filled in and under the Technical Processing Statement the word main has been entered. This is the minimum amount of information that can presently be entered to produce card sets, spine labels, and book pocket labels.

Figure 2 shows a request for cards and book pocket labels for 2 copies, one of which will be in the reference collection. The library has also requested one extra main entry card. Note that the L on the last line has been circled. This is done to indicate suppression of spine labels.

Figure 3 is an example of a request for catalog cards and spine labels for three copies, copies one and two are for the main library and copy three is for the Rhode Island collection shelf location. (c.3 R.I. CL) The library has also filled in the local call number line (CN□). This number will appear on all catalog cards and spine labels for this order. The P is circled to prevent pocket label production.

Figure 4 shows a request for catalog cards, book spine labels, and book pocket labels for three books, v.1 c.1 to the central library card v.1 c.2, v.2 c.2 to the NML branch library. Two groups of catalog cards will be produced for this request: group one will use the general library formats and group two will use the branch library formula. The spine labels and the margin area of the catalog cards for the branch (NML) request will be headed with the branch abbreviation in this case NML.

Figure 5 is a sample request for catalog cards and book spine labels for eight items consisting of copies 1-3 of volumes 2 and 3 for the central library, and copies 1-2 of volume 1 for the branch library called EXT. The library is also requesting that the local call number be used in margin of the card rather than the normally supplied L.C. call number. Two extra main entry cards have been requested and book pocket labels are not to be produced.

The following is an image of what each spine label would look like from this request (Figure 5).

| | | | | | | | |
|------|------|------|------|------|------|-----|-----|
| 824 | 824 | 324 | 824 | 324 | 824 | EXT | EXT |
| .8 | .8 | .3 | .8 | .3 | .8 | 824 | 824 |
| 1964 | 1964 | 1964 | 1964 | 1964 | 1964 | .8 | .8 |
| v.2 | v.2 | v.2 | v.3 | v.3 | v.3 | v.1 | v.1 |
| c.1 | c.2 | c.3 | c.1 | c.2 | c.3 | c.1 | c.2 |

NELINET Request Form

RI

LC ☐ 63-17953
L.C. Card Number

| AT <input type="checkbox"/> | Location | Copy-Shelf Statement | Volume |
|-----------------------------|----------|----------------------|--------|
| AT <input type="checkbox"/> | | | |
| AT <input type="checkbox"/> | | | |
| AT <input type="checkbox"/> | | | |

| | <u>Branches</u> | <u>Shelf Locations</u> |
|---|-----------------|------------------------|
| CN <input type="checkbox"/> <u>Local Call Number</u> | NML. | Rare |
| | GLS | R.I.Cl |
| XC <input type="checkbox"/> <u>Extra Main Entry Cards</u> | EXT | J.F.K. |
| | | Blatz |
| | | Ref |
| N <u>CLP</u> | | Juv. |
| <u>No Cards Labels Pockets</u> | | Archiv |
| | | Thesis |
| | | mfilm |
| | | mc card |
| | | mfiche |

FIGURE 1

NELINET Request Form

RI

LC ☐ 63-17953
L.C. Card Number

| AT <input type="checkbox"/> | Location | Copy-Shelf Statement | Volume |
|-----------------------------|----------|----------------------|--------|
| | MAIN | ; C.1, C.2 Ref | |
| AT <input type="checkbox"/> | | | |
| AT <input type="checkbox"/> | | | |
| AT <input type="checkbox"/> | | | |

CN ☐
Local Call Number

XC ☐ 1
Extra Main Entry Cards

N C L P
No Cards Labels Pockets

Branches

NML
GLS
EXT

Shelf Locations

Rare
R.I.C1
J.F.K.
Blatz
Ref
Juv.
Archiv
Thesis
mfilm
mcard
mfiche

FIGURE 2

NELINET Request Form

| | | | |
|--------------------------|------------------|---------------------------|--------|
| <input type="checkbox"/> | <u>63-23627</u> | | |
| | L.C. Card Number | | |
| <input type="checkbox"/> | <u>MAIN ;</u> | <u>1C.1-2, C.3 R.I.CL</u> | |
| | Location | Copy-Shelf Statement | Volume |
| <input type="checkbox"/> | | | |
| <input type="checkbox"/> | | | |
| <input type="checkbox"/> | | | |
| <input type="checkbox"/> | | | |

| | | | |
|--------------------------|--------------------|--|--|
| <input type="checkbox"/> | <u>574/. 92074</u> | | |
| | Local Call Number | | |

| | | | |
|--------------------------|-------------------------------|-----------------|------------------------|
| <input type="checkbox"/> | <u>Extra Main Entry Cards</u> | <u>Branches</u> | <u>Shelf Locations</u> |
| | | NML | Rare |
| | | GLS | R.I.C1 |
| <input type="checkbox"/> | <u>C L (P)</u> | EXT | J.F.K. |
| | No Cards Labels Pockets | | Blatz |
| | | | Ref |
| | | | Juv. |
| | | | Archiv |
| | | | Thesis |
| | | | mfilm |
| | | | mc card |
| | | | mfiche |

FIGURE 3

NELINET Request Form

RI

LC ☐64-11130
L.C. Card NumberAT ☐MAIN ; C. 1
LocationCopy-Shelf Statement ; V. 2
VolumeAT ☐NML ; C. 2; V. 1-2AT ☐AT ☐CN ☐Local Call NumberBranchesShelf LocationsXC ☐Extra Main Entry CardsNML
GLS
EXTRare
R.I.Cl
J.F.K.
Blatz
Ref
Juv.
Archiv
Thesis
mfilm
mcard
mfiche

N

C L P
No Cards Labels Pockets

FIGURE 4

NELINET Request Form

RI

LC ☐ 64-14237
☐ L.C. Card Number

AT ☐ MAIN ; 1 C. 1-3 ; 1 V. 2-3
☐ Location Copy-Shelf Statement Volume

AT ☐ EXT ; 1 C. 1-2 ; 1 V. 1
☐ Location Copy-Shelf Statement Volume

AT ☐ _____
☐ Location Copy-Shelf Statement Volume

AT ☐ _____
☐ Location Copy-Shelf Statement Volume

CN ☐ 8241.8 / 1964
☐ Local Call Number

BranchesShelf Locations

XC ☐ 2
☐ Extra Main Entry Cards

NML
 GLS
 EXT

Rare
 R.I.Cl
 J.F.K.
 Blatz
 Ref
 Juv.
 Archiv
 Thesis
 mfilm
 mcard
 mfiche

N C L (P)
☐ No Cards Labels Pockets

FIGURE 5

G - 1

APPENDIX G

COMPRESSION WORD CODING TECHNIQUES:
TRANSITION DISTANCE CODING, ALPHACHECK
RECURSIVE DECOMPOSITION, AND SOUNDEX.

"I have said enough to convince you that ciphers of this nature are readily soluble, and to give you some insight into the rationale of the development. But be assured that the specimen before us appertains to the very simplest species of cryptograph."

--Edgar Allen Poe
The Gold Bug

G1. INTRODUCTION

Cryptographic studies have documented much useful language data having application to retrieval coding. Because unclassified cryptographic studies are few, Fletcher Pratt's 1939 work¹ remains the classic in its field. Gaines² has the virtue of being in print, and the more recent cryptographic history of Kahn³, while comprehensive, lacks the statistical data that made the earlier works valuable. The work coding problem for language processing, as opposed to cryptography, has been extensively studied by Nugent and Vegh⁴. Information theorists have contributed the greatest volume of literature on coding, and have added to its mathematical basis, largely from the standpoint of communications and error avoidance.

We present here a brief discussion of compression codes and their objectives, and then describe four compression codes having application to retrieval directories.

Transition Distance Coding is a new method that has been devised for this project. It is a randomizing code that results in short codes of high resolving power.

Alphacheck has also been devised for this project, and combines high readability with good resolution. It permits simple truncation to be used by means of applying a randomized check character that acts as a surrogate of the omitted portion. It appears to have the greatest potential, in directory applications, of the codes considered here.

Recursive Decomposition is a selected letter code that was devised by the author several years ago⁴. It has been tested and has the advantages of simple derivation and high resolution.

Soundex⁵ is the only compression code that has achieved wide usage. It was devised at Remington Rand for name matching under conditions of uncertain spelling.

G2. OBJECTIVES OF COMPRESSION CODING

It is desired to transform sets of variable length words into fixed length codes that will maximally preserve word to word discrimination. In the final directories to be used, the codes for several elements will be accessible to enable the matching of several factors before a file record is selected. The separate codes for differing factors need not be the same length, though each type of code will be of uniform length, nor need the codes for differing factors be derived by the same process.

What we loosely call codes, must be formally ciphers. That is, they must be derivable from the data words themselves, and not require "code books" to determine equivalences. This is so because the file directories must be derivable from file items,

entries in directory form must be derivable from an input query, and these two directory items must match when a record is to be extracted. The ciphers need not be decipherable for our application, and in general are not.

Fixed length is generally desirable for machine directories since this provides the rough equivalent and simplicity of a margin entry in a paper directory. We will examine the question of variable length directories and directory list structures in a later memo.

The functions of the codes will determine their form, and a code or file key designed to meet one objective will generally not be satisfactory for any other objective. We will clarify this point by illustrating some typical objectives:

- (a) File key for extraction of records in approximate file order. (Sorting and Print-out Problem.)

Typical code construction rule: Take first six letters.

JOHNSEN \Rightarrow JOHNSE
JOHNSON \Rightarrow JOHNSO
JOHNSTON \Rightarrow JOHNST
JOHNSTONE \Rightarrow JOHNST

- (b) File key for extraction of records under conditions of uncertainty of spelling (Air-line reservation problem.)

Typical code construction rule: Vowel Elimination or Soundex.

Typical matching rule: Best match.

Vowel Elimination

JOHNSEN \Rightarrow JHNSN
JOHNSON \Rightarrow JHNSN
JOHNSTON \Rightarrow JHNSTN
JOHNSTONE \Rightarrow JHNSTN

Soundex

J525 \Rightarrow J52
J525 \Rightarrow J52
J5235 \Rightarrow J52
J5235 \Rightarrow J52

- (c) File key extraction of records from accurate input, with objective of maximum discrimination of similar entries (cataloging search problem.)

Typical code construction rules: Recursive Decomposition Coding or Transition Distance Coding.

Recursive DecompositionTransition Distance

| | | |
|--------------|---------|------|
| JOHNSEN == | JHNSEN | EFTZ |
| JOHNSON == | JHNSON | DNWU |
| JOHNSTON == | JHSTON | ZIKY |
| JOHNSTONE == | JHSTONE | ECRC |

In file keys we are concerned with presently, we assume accurate input data and the objective is maximum discrimination. Since it would be nice to have our cake and eat it too, we would like a code to be as discriminating as Transition Distance Coding and to be as readable as truncation coding. We achieve this, possibly, by combining the two codes into one, with an initial portion truncated and a final check character representing the remainder via a compressed Transition Distance Code: Alphacheck.

- (d) File key for human readability and high word to word discrimination.

Possible code construction rules: Alphacheck:
Simple truncation plus a terminal check character.

| | |
|---------------|--------|
| JOHNSEN ==> | JOHNSV |
| JOHNSON ==> | JOHNSX |
| JOHNSTON ==> | JOHNSD |
| JOHNSTONE ==> | JOHNS3 |

We describe these procedures in the following sections.

G3. TRANSITION DISTANCE CODING

It is axiomatic that randomizing codes give the greatest possible discrimination for a given code space. The whole trick of creating a good compression code is to eliminate the natural redundancy of English orthography, and preserve discrimination in a smaller word size.

Letter-selection codes can only half accomplish this, due to the skewed distribution of letter usage. They can eliminate the higher frequency components, but they cannot increase the use of the lower frequency components.

Randomizing codes - often called "hash" codes, properly quasi-random codes - can equalize letter usage and hence make best

use of the code space. Prime examples here are the variants of Gödel coding devised by Vegh⁴ in which the principle of obtaining uniqueness via the products of unrepeatd primes is exploited, as it is in the randomizing codes we consider here. The problem in design of a randomizing code, is that the results can be skewed rather than uniformly distributed due to the skewed nature of the letters and letter sequences that the codes operate on.

In Transition Distance Coding, we overcome the natural bias of letters and letter sequences, by operating on a word parameter that is itself semi random in nature. We advance the following principle, not quite a theorem.

Principle: Considering letters in their normal ordinal alphabetic position, and considering letter transitions to be unidirectional and cyclic, the distribution of transition distances in English words is essentially uniform.

In view of the fact that letter usage has an extremely skewed distribution, with a probability ratio in excess of 170 to 1 for the extremes, it is seen that the more uniform parameter of transition distances is a superior one for achieving randomized codes. The relative uniformity of transition distance needs further investigation, but one typical letter diagram sample from Gaines² with 9999 transitions (means number of occurrences of each distance = 335) yielded a mean deviation of 99 and a standard deviation of 123, and an extreme probability ratio of 3.3 to 1 for the different transition distances from 0 to 25. The distribution can be made more uniform by letter permutation. Permutation is used in the algorithm for Transition Distance Coding but not in Alphacheck. It's value will be determined.

G3.1 ALGORITHM

The method of Transition Distance Coding is used to operate on a variable length word to achieve fixed length alphabetic or alphanumeric codes that exhibit quasi-random properties. The code is formed from the modulo product of primes associated with transition distances of permuted letters. The method is intended strictly for computer operation, as it is a simple program but an extremely tedious manual operation. There are five steps:

- (1) Permute characters of natural language word. This breaks the diagram dependency that could make the transition distances less uniformly distributed. This step might be dispensed with if the resulting distributions prove satisfactory without it. The permutation process consists of

taking the middle letter (or letter right of middle for words with an even number of letters), the first, the last, the second, the next-to-last, etc. until all letters have been used. That is, for a letter sequence:

$$a_1, a_2, \dots, a_i, \dots, a_n$$

The following permutation is taken:

$$a_{\text{Int}(\frac{n+1}{2})}, a_1, a_n, a_2, a_{n-1}, \dots, a_{(1+i)}, a_{(n-i)}, \dots, a_{\text{Int}(\frac{n+1}{2}) + \text{Rem}(\frac{n}{2})}$$

where Int and Rem refer to the integer part and remainder, respectively. To illustrate a typical case:

JOHNSEN \Rightarrow N J N O E H S

- (2) Take transition distances of the characters. We assign letters a position value corresponding to their normal ordinal alphabetic positions excepting Z, which we equate to 0, (e.g., A = 1, Y = 25, Z = 0), and take the transition distances between successive letters of the input sequence. Distance is measured unidirectionally in alphabetic order, and cyclicly (i.e., "around the bend," Z to A.) The sequence AX has the transition distance $N_X - N_A = 24 - 1 = 23$. Negative distances are converted to their positive cyclic distance by taking their 26 complement. That is, the sequence XB has the transition distance $N_B - N_X = 2 - 24 = -22 \Rightarrow 26 - 22 = 4$. To follow the 'JOHNSEN' example:

$$\text{NJNOEHS} \Rightarrow (14, 10, 14, 15, 5, 8, 19) \Rightarrow \begin{matrix} \text{letter} & \text{distances} \\ \text{numbers} \end{matrix} (22, 4, 1, 16, 3, 11)$$

- (3) Associate with each transition distance, a corresponding prime number. Table G1 shows the primes corresponding to the transition distances. We begin with 5 so that the alphanumeric base (36) and all numbers are relatively prime. We follow the example above:

$$(22, 4, 1, 11, 3, 11) \Rightarrow \begin{matrix} \text{distances} & \text{primes} \end{matrix} (39, 13, 5, 61, 11, 41)$$

| <u>LETTER</u> | <u>LETTER POSITION AND DISTANCE VALUE</u> | <u>PRIME NUMBER</u> |
|---------------|---|---------------------|
| A | 1 | 5 |
| B | 2 | 7 |
| C | 3 | 11 |
| D | 4 | 13 |
| E | 5 | 17 |
| F | 6 | 19 |
| G | 7 | 23 |
| H | 8 | 29 |
| I | 9 | 31 |
| J | 10 | 37 |
| K | 11 | 41 |
| L | 12 | 43 |
| M | 13 | 47 |
| N | 14 | 53 |
| O | 15 | 59 |
| P | 16 | 61 |
| Q | 17 | 67 |
| R | 18 | 71 |
| S | 19 | 73 |
| T | 20 | 79 |
| U | 21 | 83 |
| V | 22 | 89 |
| W | 23 | 97 |
| X | 24 | 101 |
| Y | 25 | 103 |
| Z | 0 | 107 |

TABLE G1

Letter Positions and Primes used in
Transition Distance Coding and Alphacheck

- (4) Multiply these primes, modulo the capacity of the computer. Integer multiplication in single precision is effected, disregarding overflow. For a computer with an 18 bit word length containing a 1 bit sign position, we multiply modulo 2^{17} . That is, we disregard product portions that equal or exceed 131,072. For a machine of this type, then, we will be generating a quasi random number in the range of 0 to 131,071. This is converted to alphanumeric form in the next step. Following the example:

$$39 \times 13 \times 5 \times 61 = (352,385) \text{ Mod } 2^{17} = 90,741$$

$$90,741 \times 11 = (998,151) \text{ Mod } 2^{17} = 80,647$$

$$80,647 \times 41 = (3,306,527) \text{ Mod } 2^{17} = 39,727$$

- (5) Convert to alphabetic or alphanumeric form. We now express the number derived above as an integer base 26 (alphabetic form) or base 36 (alphanumeric form). We will use a 4 digit code. In the case of alphabetic representation we use the letters to represent the numbers of their ordinal position (A=1, B=2, etc.), and use Z as zero. In alphanumeric form we would use the digits 0 to 9 to represent this range, and the letters A through Z would represent the range from 10 to 35.

Using the 18 bit word length we have assumed, the alphabetic form is as good as the alphanumeric. The range of the random number extends to 131,071; the range of 4 digit alphabetic representation extends to $(26^4 - 1) = 256,975$; the range of 4 digit alphanumeric representation extends to $(36^4 - 1) = 1,676,615$. Hence, the alphabetic representation is sufficient. We divide the random number successively by 26^3 , 26^2 , 26^1 , and 26^0 to obtain the alphabetic form. We follow the example:

$$39,727 / 26^3 = 2 + \text{Rem } 4575 \Rightarrow B$$

$$4575 / 26^2 = 6 + \text{Rem } 520 \Rightarrow F$$

$$520 / 26^1 = 20 + \text{Rem } 0 \Rightarrow T$$

$$0 / 26^0 = 0 \Rightarrow Z$$

$$\text{JOHNSEN} \Rightarrow \text{BFTZ}$$

G4. ALPHACHECK

Alphacheck is a means for creating a randomized alphanumeric check digit. When used with a selected letter compression code, it operates on the missing letters to generate a single character surrogate. We use it to add discrimination to a simple truncation code, and thereby we hope to attain a compression code that is both readable and resolving.

A process practically identical to that of Transition Distance Coding is used, except that at the final step the random number is taken modulo 36 and expressed as an alphanumeric character. The ten numeric digits represent themselves, and the letters A to Z represent the mod 36 numbers from 10 to 35, or their ordinal alphabetic value plus 9.

In this case, the difference between an alphabetic representation and an alphanumeric one is significant, since only one character is used, and the range of the Alphacheck character is much smaller than the range of the binary random number it is derived from.

The probability of no repetition of Alphacheck codes in a sample of size r , is a case of determining the probability of uniqueness for sampling with replacement from a population n , for which:

$$p = \frac{n!}{n^r (n-r)!}$$

where n is the range of the code, for alphanumeric Alphacheck, $n = 36$.

The median of the distribution of p , r_m gives the sample size for which the probability of uniqueness is 0.5. This is estimated by taking the logarithmic form of p , which yields a good approximation when n is large with respect to r .

$$\ln p \approx - \frac{r^2}{2n}$$

$$r_m \approx [2n \ln(.5)]^{1/2} = 1.18 n^{1/2} = 7.08$$

By comparison, r_m for $n=26$ is 6.05; for $n=131,072$ (Transition Distance Coding in 4 characters and modulo 2^{17}) r_m is 427.

We may conclude that the alphanumeric Alphacheck (36 symbol) has a 50% expectation of uniquely resolving 7 otherwise identical 5-letter truncations of source words; this is a one word advantage over the 26 symbol alphabetic Alphacheck. Hence, we will use the alphanumeric form.

G4.1 ALGORITHM

It is not appropriate to use the identical randomizing method of T.D.C. (Transition Distance Coding), since this was designed to operate on full words, whereas we wish to operate on the omitted remainders of truncated words, which are often as short as two letters. When a two letter remainder exists only one transition distance is involved, and hence only one prime number; and the individual primes are not uniformly distributed modulo 36. Hence, in the case where only one transition distance exists, the corresponding prime is multiplied by two additional primes corresponding to the letters involved (Table G1.) If only two distances are involved, we associate another prime corresponding to the last letter. Since randomization is created largely by the multiplicative properties of the process, we insure that at least three factors are multiplied in all cases. Except for this difference in step three, the randomizing process is essentially identical to that of TDC. The steps are:

- (1) If word is six letters or less take whole word, otherwise, take first five letters and compute an Alphacheck character for the sixth, based on the omitted letters.
- (2) Take transition distances of the omitted letters (as in TDC).
- (3) Associate with each transition distance a corresponding prime number (as in TDC). If only one transition distance exists, additionally associate prime numbers with the remaining letters. If only two transition distances exist, additionally associate a prime number with the last letter.
- (4) Multiply these primes, modulo the capacity of the computer (as in TDC).
- (5) Convert to alphanumeric form in 1 symbol, modulo 36, in which $0 \Rightarrow 1$, \dots , $9 \Rightarrow 9$, $10 \Rightarrow A$, $11 \Rightarrow B$, \dots , $35 \Rightarrow Z$.

The example of the JOHNS - names, shown in Table G2 illustrates the process.

| Name | JOHNSEN | JOHNSON | JOHNSTON | JOHNSTONE |
|----------------------|---------|---------|----------|------------|
| Truncated Portion | JOHNS | JOHNS | JOHNS | JOHNS |
| Remainder | EN | ON | TON | TONE |
| Letter # | 5,14 | 15,14 | 20,15,14 | 20,15,14,5 |
| Distance | 9 | 25 | 21,25 | 21,25,17 |
| Distance Primes | 31 | 103 | 33,103 | 33,103,67 |
| Letter Primes | 17,53 | 59,53 | 53 | - |
| Product | 27,931 | 322,081 | 453,097 | 572,783 |
| Mod 2 ¹⁷ | 27,931 | 59,937 | 59,331 | 48,495 |
| Mod 36 | 31 | 33 | 13 | 3 |
| Alphacheck Character | V | X | D | 3 |
| Resulting Code | JOHNSV | JOHNSX | JOHNSD | JOHNS3 |

TABLE G2

Example of Key Generation by Alphacheck

G5. RECURSIVE DECOMPOSITION CODING

This method uses a frequency ordering of letters, and selection or rejection of a particular letter is based on that letter's relative order in the table with respect to the previous letter. It thus gives a statistical advantage, though not an absolute one, to the lower frequency letters. Since many words differ only in high frequency vowels (e.g., COMPUTE, COMPETE, COMPOTE), this relative order feature adds a randomizing aspect to selection that permits inclusion of occasional high frequency letters.

The frequency ordering used is taken from tables in Pratt¹. Different word samples will yield slightly different orderings, but the cipher resolution is not sensitive to minor orderings. The Pratt ordering is:

E T A O N R I S H D L F C M U G Y P W B V K X J Q Z

The algorithm is:

"If a source word is longer than six letters, select the first letter and subsequent letters of lesser or equal ordering than the prior letter, and continue the process recursively until six letters remain. Words of six letters or less are reproduced in full and filled out with null symbols, where necessary, until a total of six characters is reached."⁴

Several examples will illustrate the system. Omitted letters are shown circled, and successive cycles are shown by arrows.

1. B I B L I O G R A P H E R \Rightarrow B B G P E R
2. I N F O R M A T I O N \Rightarrow I F R M I N
3. S H A K E S P E A R E \Rightarrow S H K S P A R \Rightarrow S H K P A R
4. S M I T H \Rightarrow S M I T H \square
5. K I N G S F O R D - S M I T H \Rightarrow K G F R D M H \Rightarrow K F R D M H
6. K R I S H N A M O O R T H I \Rightarrow K I S H M O R H \Rightarrow K S H M R H

In some very rare cases, an emerging cipher may have more than six letters in descending sequence, so that it will not decompose further. In such cases the final letters are eliminated until six remain.

Most words, however, will reduce in one or two cycles. In a test of 55,000 words only one was found requiring four cycles. A few extreme cases do exist, however: the longest ever found required six cycles:

7. A N T I D I S E S T A B L I S H M E N T A R I A N I S M ⇒
 A N I D S S A B S H M N A R I N I S M ⇒
 A N I D S B H M R I I S M ⇒
 A N I D B M I I S M ⇒
 A N I D B I S M
 A N I D B S M
 A N I D B M

Only slightly shorter, the longest word in Shakespeare's works (Love's Labour's Lost V.i) reduces in three recursions:

8. H O N O R I F I C A B I L I T U D I N I T A T I B U S ⇒
 H N R I F C B L U I A I B ⇒
 H R I F C B U I B ⇒
 H I F C B B

Even Mary Poppins' sesquipedalian ecphronesis crumbles to six letters in three recursions:

9. S U P E R C A L I F R A G I L I S T I C E X P I A L I
 D O C O U S ⇒
 S U P R C L F G L S I C X L D C U ⇒
 S U P C F G C X C U ⇒
 S U P G X U

The prime advantages of the method are its computational simplicity and its resolution. The elimination requires only table lookup and no multiplications; and the compression is readily done manually. The resolution is apparently as good as one can get with a selected letter compression code. It effectively flattens the high portions of the letter frequency curve, though unlike a randomizing code, it cannot totally equalize the distribution. The resolution, however, is quite good. Specifically in a test of 1362 words (chosen from the secretary's handbook 20,000 Words), only thirty of the six letter ciphers (about 0.61%) were non-unique

and of the non-unique ciphers all were simple pairs except for one instance of three occurrences. The method compresses quickly: since all non-initial letters have a .5 probability of being retained, the expected length, L , of an n letter word after r recursions is:

$$L = 1 + \frac{n-1}{2^r}$$

This indicates that a 43 letter word may be expected to compress to six letters in three recursions.

G6. THE SOUNDEX CODE

The Soundex code⁵, though widely used, is of obscure origin and has been attributed to Remington Rand. It is a phonetic code that tends to create identical codes from similar sounding names. It is useful for name searching under conditions of uncertainty of spelling, such as occurs in the airline reservation problem where it is often required to match a telephoned name in a machine file. The code has five steps:

1. Retain first letter of name as first letter of code.
2. Eliminate vowels, W, H, and Y.
3. Eliminate the second consonant of a double consonant pair.
4. Replace the following letters by numbers:

| | |
|------------------------------|---|
| B,P,F,V | 1 |
| C,G,J,K,Q,S,X,Z,SC,CH,SCH,CK | 2 |
| D,T | 3 |
| L | 4 |
| M,N | 5 |
| R | 6 |

5. Take the first three or four symbols, and add zeros if insufficient phonetic sounds.

The example below illustrates the process:

JOHNSEN \Rightarrow JNSN \Rightarrow J525 \Rightarrow J52
 JOHNSON \Rightarrow JNSN \Rightarrow J525 \Rightarrow J52
 JOHNSTON \Rightarrow JNSTN \Rightarrow J5235 \Rightarrow J52
 JOHNSTONE \Rightarrow JNSTN \Rightarrow J5235 \Rightarrow J52

REFERENCES

1. Pratt, Fletcher Secret and Urgent, the Story of Codes and Ciphers Blue Ribbon Books, Garden City, New York 1939.
2. Gaines, Helen Fouche' Cryptanalysis Dover, New York, 1956.
3. Kahn, David The Codebreakers Macmillan Company, New York, 1967.
4. Nugent, William R. and Vegh, Alexander Automatic Word Coding Techniques for Computer Language Processing Report RADC-TDR-62-13 February, 1962 Volume I: AD-272-401, Volume II: AD-272-402.
5. Wright, M.A. "Mechanizing a Large Index; Appendix: The Soundex Code" The Computer Journal v.3, July, 1960 p.33.

APPENDIX H: SAMPLE CARDS AND LABELS

SELIN LABELS
(for spine
of book)

PRESSURE
SENSITIVE
LABELS
(for book
pockets)

HD
38
.G725

Greenwood, William T., comp.
Business policy; a management audit
MANAGEMENT AUDIT--CASE STUDIES.

LB
1051
.S7239

KF
9625
.T5

LB
1576
.T56

B
825
.P57
1967

HD
38
.G725

Greenwood, William T., comp.
Business policy; a management audit
Business policy

HD
5764
.A6
K5

HD
38
.G725

Greenwood, William T., comp.
Business policy; a management audit

PQ
4835
.17
2539

HD
38

Greenwood, William T., comp.
Business policy; a management audit
.G725

HD
38
.G725

Greenwood, William T., comp.
Business policy; a management audit

HD
38
.G725

Greenwood, William T., comp.
Business policy; a management audit
approach, by William T. Greenwood.
New York, Macmillan [1967]
xiv, 656 p. illus., maps. 24 cm.
Includes bibliographical references.
1. Management audit--Case studies.
I. Title.

67-14417
658

HD
38
.G725

Greenwood, William T., comp.
Business policy; a management audit
approach, by William T. Greenwood.
New York, Macmillan [1967]
xiv, 656 p. illus., maps. 24 cm.
Includes bibliographical references.
1. Management audit--Case studies.
I. Title.

This library requires
four main entries.

n.h.
HD38.G725

67-14417
658

T 15 Technology in western civilization.
Edited by Melvin Kranzberg [and]

TECHNOLOGY AND CIVILIZATION.

T 15 Technology in western civilization.
Edited by Melvin Kranzberg [and]

TECHNOLOGY--HISTORY.

T 15 Technology in western civilization.
Edited by Melvin Kranzberg [and]

Pursell, Carroll W., ed.

T 15 Technology in western civilization.
Edited by Melvin Kranzberg [and]

Kranzberg, Melvin, ed.

T 15 Technology in western civilization.
Edited by Melvin Kranzberg [and]
.T43 Carroll W. Pursell, Jr. Executive

T 15 Technology in western civilization.
Edited by Melvin Kranzberg [and]
.T43 Carroll W. Pursell, Jr. Executive
editors: Paul J. Grogan [and] Donald
F. Kaiser. New York, Oxford
University Press, 1967-
v. illus., maps, ports. 24 cm.
Bibliography: v.1, p. 745-774.
Contents.--v. 1. The emergence of
modern industrial society, earliest
times to 1900.

(Cont. next card)

67-15129

T 15 Technology in western civilization.
1967 (Card 2)

TECHNOLOGY AND CIVILIZATION.

T 15 Technology in western civilization.
1967 (Card 2)

TECHNOLOGY--HISTORY.

T 15 Technology in western civilization.
1967 (Card 2)

Pursell, Carroll W., ed.

T 15 Technology in western civilization.
1967 (Card 2)

Kranzberg, Melvin, ed.

T 15 Technology in western civilization.
1967 (Card 2)

T 15 Technology in western civilization.
1967 (Card 2)

.T43

1. Technology--Hist. 2. Technology
and civilization. I. Kranzberg,
Melvin, ed. II. Pursell, Carroll W.,
ed.

n.h.
T15.T43

67-151